

Room for hope after a chemical defeat

The Clinton administration blinked and a small group of arch-conservatives dealt an unexpected blow to the Chemical Weapons Convention. The treaty has its weaknesses, but it merits greater faith in its strengths.

If ever there was an object lesson in the perils of complacency, last week's Republican snatching of defeat from the jaws of a Clinton victory has provided it. A decision to postpone ratification by the US Senate of the Chemical Weapons Convention (CWC) could have been avoided had the arch-conservative Jesse Helms not succeeded in delaying its consideration last year: if hard-line Republican activists had thus been prevented from developing an anti-treaty lobby amongst small chemical companies; and if Republicans in the Senate had not then been able surreptitiously to use it to persuade their colleagues, and their leader, Robert Dole, that here was a handy way to differentiate themselves from the Democrats during the run-up to an election.

That, rather than deep reconsiderations of the rights and wrongs of the convention, is what allowed Republican senators to carry the day last week. No-one is defending possible military deployment or trade in nerve agents such as sarin, soman and tabun, mustard agents, hydrogen cyanide, and biochemical toxins. No opposition has been voiced to the requirement of the convention that an international agency be set up in the Hague to implement the treaty's provisions; that signatories should declare their chemical weapons stocks and plans to destroy them within ten years of the treaty coming into force, and to close production facilities; or that signatories should be open to inspection.

The electoral tactics were successful because opponents have been able to attack the treaty's soft underbelly. Companies, they say, could find themselves forced to be open to inspection at the whim of other signatory countries. To make matters worse, the inspectors might come, for example, from two of the nations least trusted by US industrialists: Japan and France (both with past episodes of industrial espionage). Then there would be all that bureaucratic form-filling insisted on by the Hague agency. And, they say, one should consider ratifying the convention only if Iraq, Iran and Libya will sign up to it.

To give these countries such a veto would be a sure way of scuppering the convention. That posture should be dismissed as the spoiling tactic that it is. But the other concerns are less easily dealt with, because there is an element of truth in them. "Suck it and see" is the flavour of the response of treaty supporters to them. In its favour, the convention is stiff with provisions ensuring that those protesting against an inspection, or aspects of it, get their hearing, and that a substantial majority of signatories then have to be convinced that the inspection is indeed called for. Strength is added to the treaty defenders' case by the fact that 63 countries have already signed up to it, including some that are hardly negligible in their chemical interests such as Germany, the United Kingdom and Sweden. Sixty-five are required for the convention to come into force.

Another problem for treaty proponents is that the role of chemicals is not always unambiguous. If a developing country is found to be harbouring large stocks of *O*-ethyl dimethylamidophosphoryl cyanide, one can be certain that it is willing to destroy the nervous systems of its opponents. In contrast, a cache of isopropanol may be, but is not necessarily, intended to be mixed with other compounds in a binary weapon to rain sarin on

the heads of enemies. All the same, a major munitions effort does tend to leave its traces, and one can set considerable store by a combination of satellite surveillance, intelligence and the on-site inspections specified in the convention.

Such considerations may all be beside the point in the United States if last week's events are a political anomaly — an obstacle riding a wave of temporary political cohesiveness that will evaporate in the convention's favour if President Bill Clinton is re-elected. Domestic considerations in the CWC's favour are strong: the influence of the United States on the treaty's implementation will be minimized if it stays out while 65 others ratify it. (But a *de facto* presence would still be felt, given the country's sheer chemical prowess and its extensive past involvement in the convention's development.)

One factor that will help to tip the scales, apart from the weakness of the opponents' case, is the fact that major chemical companies stand to lose substantial exports, following embargoes that can be applied to non-signatories under the terms of the convention five years after it comes into force. But nevertheless, if re-elected, Clinton and his administration will need to concentrate much harder the next time the CWC surfaces in the Senate.

BSE's propaganda

The UK government's public deployment of a *Nature* paper encourages incorrect perceptions of the role of science.

CLUTCHING a copy of R. M. Anderson *et al.* (*Nature* **382**, 779–788; 1996) the British government this week asked the European Commission to reduce a cull of 127,000 cattle demanded by the commission as part of plans to reduce the incidence of bovine spongiform encephalopathy (BSE) in the UK herd. The paper's epidemiological analysis of the BSE epidemic suggests that this will cease around 2001 even without culling, and that no culling strategy short of shooting almost 3 million cattle would make a significant dent in the epidemic before then (see page 209). According to the government (in public, at least), this justifies doing nothing.

Britain's European neighbours are not so sanguine. The UK government's credibility in the eyes of the continental European public could hardly be lower. And what surprises Britain's neighbours most is why Britain is making such a fuss about the cull. And they argue that, according to the same analysis, a cull of 150,000 cows could eliminate more than 2,000 (or about a third) of the total cases predicted to occur before the end of the epidemic. Britain has culled similar numbers during other epidemics. To other Europeans, arguing about the rationality of the public's response or the scientific grounds for the cull seems neither here nor there.

The role of science, with its attendant uncertainties, is to illuminate political choices, not to enforce them. By acting as if it is oblivious to this truth, and to European political reality, the UK government can only erode its credibility even further.



NASA head fends off proposal to bring in 'base closure' scheme

Washington. The US National Aeronautics and Space Administration (NASA) is unlikely to meet its ambitious goals for cutting back on facilities and personnel to meet future budget predictions, due partly to internal 'turf battles' and partly to outside political pressure, according to a report released last week by the congressional General Accounting Office (GAO).

In fact, says GAO, NASA may eventually need an independent base-closure commission, similar to one that reviewed military bases, to help it to decide where politically sensitive cuts should be made.

The report points out that NASA's efforts to cut \$4 billion — or 25 per cent of its facility costs — by 2000, and to trim its

21,500-person workforce by nearly 20 per cent, have run into a number of problems. First is the agency's own problem in accurately predicting cost reductions. In some cases, NASA has overestimated how much money could be saved by a certain action, such as moving all the agency's research aircraft to California's Dryden Flight Research Center (see *Nature* 382, 745; 1996).

Plans to consolidate telecommunications and supercomputing capabilities have suffered from flawed or incomplete analysis. And attempts to find savings by cost-sharing with the Defense Department (DOD) have met with little success: "Each NASA and DOD centre or laboratory wants to protect its ability to maintain technical expertise and

competence," says GAO.

More fundamental, however, are the political problems of 'downsizing'. A proposal from NASA earlier this year to reduce dramatically the number of personnel at its headquarters received an angry response from members of Congress in the Washington area, and the cuts are now expected to be less severe. The plan to shift aircraft to Dryden prompted congressional delegations representing other NASA centres to write to the NASA administrator, Dan Goldin, earlier this month to "express their concern".

According to the GAO report, NASA's own efforts to close facilities or relocate activities have "been slowed by parochial concerns about the effects of such actions on missions, personnel, and local communities". As a result, it says, NASA "may not be able to successfully downsize infrastructure, especially facilities, on its own".

A neutral closure commission has therefore been proposed, which would provide political cover for any hard decisions. It was NASA and DOD officials who suggested the idea to GAO, according to Thomas Schulz, the study's lead auditor.

But Goldin, who has built a reputation as a reformer, last week told the House of Representatives government reform committee that he would rather handle the problem internally. A closure commission, he said, would "destroy NASA".

Congress would have to decide whether such an extreme measure as a closure commission is necessary. The White House science office is expected to come out with its own progress report on federal laboratory downsizing next month, updating a report released last year. That report, say White House officials, is likely to be more optimistic about how much progress NASA has made in its downsizing.

If the space agency is struggling with cost reductions, the White House should bear some of the blame. NASA — like other federal agencies — is uncertain how much money it will have to spend in 2000. The Clinton administration, trying to show its budget-cutting zeal to the Republican-controlled Congress earlier this year, released projections showing NASA's budget falling from just under \$14 billion in 1997 to \$11.6 billion in 2000. But Goldin (and others in the administration) have repeatedly said those figures are not firm. As a result, NASA has no idea what budgetary target it needs to hit, and therefore how much it needs to cut.

Tony Reichhardt

Paris plan prefers art to anthropology

Paris. Art scored a victory over science last week when a commission set up by Jacques Chirac, the French president, recommended that the world's largest anthropology museum, the Musée de l'Homme (Museum of Mankind) in Paris, should be turned into a museum of primitive art. In doing so, it rejected a alternative proposal from researchers to renovate the existing museum and create a 'Grand Musée d'Histoire Naturelle de l'Homme'.

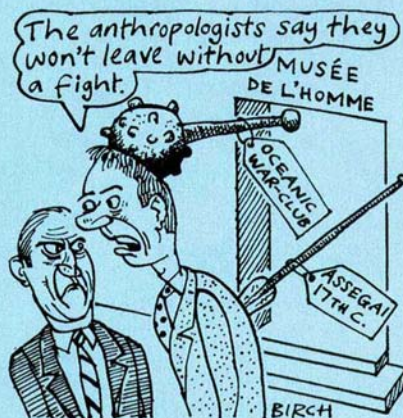
The latter proposal, which was put forward by Henry de Lumley, the director of the museum's parent body, the Natural History Museum (NHM), and a member of the commission, would draw together research on man's social, cultural and biological evolution and provide imaginative presentations of the museum's 735,000 objects (see *Nature* 383, 109; 1996).

But at a press conference in Paris last Friday, 13 September, the 'commission of early art', set up by Chirac earlier this year, recommended that the museum's collections merge with those of the Museum of African and Oceanic Arts within a new 'Museum of Mankind, Art and Civilization'. The project is estimated to cost FF1 billion (US\$200 million) over five years, according to Jacques Friedmann, the chairman of the commission and chief executive office of UAP, a French insurance company.

Under the scheme, the Natural History Museum would lose control of the reorganized museum. Moreover, the

latter would no longer be under only the ministry of education and research, but would also be the responsibility of the Ministry of Culture. The NHM would also lose its laboratory of ethnology to the new institution.

Friedmann says that most members of the commission felt that the concept of the new museum breaks down the barrier between art and science. But although science is mentioned occasionally in the scheme, what it proposes amounts essentially to a museum of 'beaux arts'. As such, it has been unanimously condemned by



museum staff, who characterize the scheme as a hostile takeover bid of a scientific museum by art specialists.

De Lumley points out that the final decision rests with Chirac, and that the museum will continue to fight for its proposal.

Declan Butler

China urged to delay 'eugenics' law ...

London. An international group of leading human geneticists has urged the Chinese government to delay implementation of its controversial 'eugenics' law until genetic legislation been discussed within the world's genetics community.

The proposal was made in a statement presented at the conclusion of the Ninth International Congress of Human Genetics in Rio de Janeiro, Brazil, last month by nine members of the congress's executive

EU aid to Palestinians focuses on hardware

Jerusalem. The European Commission has decided to focus its future aid to Palestinian universities, agreed following the signing of the Oslo peace agreement in 1993, on laboratory equipment and library books, rather than on providing general operating expenses.

According to Matthias Burchard, who oversees the distribution of the ECU41.33 million (US\$50.3 million) that the commission has provided over the past three years to the 8 universities and 20 colleges in the West Bank and Gaza Strip, the reason for the shift in focus is to help the universities to provide "better and more appropriate" higher education. "It's not a good investment to pay running costs," he explains.

The European Commission had agreed to help the universities after Arab governments, angered by the Palestinian Liberation Organization's support for Iraq during the Gulf War, cut off aid to the Palestinians. The initial three-year commitment runs out in December.

Hanan Mikhail-Ashrawi, the Palestinian Authority's Minister of Higher Education, says she agrees in principle with the commission's plan to shift the focus of investment to the universities' research infrastructure. But she also argues that travel restrictions imposed by Israel have made it impossible for the universities to operate normally and to institute necessary reforms.

"Our best-laid plans cannot be implemented because of the closure," says Ashrawi, pointing out that the Palestinian higher education system cannot manage without continued subsidies of its operating budget, and that it already lacks sufficient funds to meet commitments up to the end of the year.

According to Burchard, the commission decided this year to allocate ECU3.7 million explicitly for upgrading laboratories and libraries, reducing its subsidy of the operating budget from ECU15 million in the previous two years by the same amount.

Haim Watzman

committee,* following discussion with researchers — including several Chinese nationals — at the meeting. Their concern was focused on China's new law on Maternal And Infant Health Care, which took effect on 1 June 1995 and is widely seen as sanctioning the practice of eugenics.

Newton E. Morton, professor of human genetics at the University of Southampton, England, and one of the signatories to the statement, says Chinese scientists at the congress argued strongly for an international response to their government's new law.

He also points out that both the British Clinical Genetics Society and the Human Genetics Society of Australasia have made public their concern about certain aspects of the law, although the American Society of Human Genetics has remained silent, on the grounds that it is a purely internal matter.

Morton says he believes that an international effort is essential to help Chinese geneticists, who "are in the awkward position of either speaking up [against the law] or remaining silent".

The statement, which was presented at the final session of the congress and received almost unanimous agreement from the participants, looks beyond the situation in China, urging that the drafting of coercive genetic legislation in other countries should also be indefinitely delayed pending discussion by geneticists.

The statement cites articles in the Chinese law that have generated particular

concern among geneticists. One states that physicians must give medical advice to couples who have been diagnosed with certain genetic diseases considered to be "inappropriate" for child-bearing, and that the two may be married only if they agree to take "long-term contraceptive measures" or to be sterilized.

After consulting both Chinese and other colleagues, the signatories drafted the statement in a form that they hope will be most likely to lead to productive discussion within China. For example, they applaud the provision in the law that forbids identification of the sex of a fetus by technical means.

But the statement also points out that "other provisions of the law conflict with the counselling principles accepted by most human geneticists", and that "genetic legislation has a tragic history".

*The signatories were Bernardo Beiguelman, Universidade Estadual de Campinas, Brazil; Pedro H. Cabello and Eduardo E. Castilla, Instituto Oswaldo Cruz, Rio de Janeiro, Brazil; Alfred Knudson, Fox Chase Cancer Center, Philadelphia, Pennsylvania, USA; Henrique Krieger, Universidade de São Paulo, Brazil; Ei Matsunaga, National Institute of Genetics, Mishima, Japan; Margareta Mikkelsen, J. F. Kennedy Institute, Glostrup, Denmark; Newton E. Morton, University of Southampton, UK; and Francisco M. Salzano, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil.

Claire O'Brien

...as British scientists register protest

London. A ballot of members of Britain's Genetical Society, one of the world's oldest genetics bodies, has delivered a majority verdict in favour of withdrawing from the International Genetics Federation (IGF), in protest at its decision to hold its next meeting in China in 1998 (see above).

But the relatively low number of returns — 100 members voted out of a total of about 1,600, with 60 approving and 40 opposing a boycott — makes it unlikely that the ballot will have settled the issue, as intended.

Since 1993, when the IGF voted to hold its next congress in Beijing, the decision has been widely criticized by some members of its affiliated societies — including the Genetical Society — concerned that attendance would signal an apparent endorsement of China's controversial Maternal and Infant Health Care Law (see *Nature* **372**, 123; 1994).

In response, the IGF requested that the congress organizing committee in

China should schedule a full discussion of eugenics. Three months after the law was passed in June 1995, the IGF's request was met (see *Nature* **377**, 7; 1995), and a eugenics symposium has since been organized by an IGF member.

But some members of the Genetical Society continue to argue that the existence of the new legislation represents a serious abuse of human rights based on genetic information about individuals, and have suggested withdrawing from the IGF in protest.

The society's committee decided to ballot its members on whether they agreed to suspend the society's affiliation. The society's newsletter containing a ballot slip went to its members in July, and the results were sent out to members last week.

The society's president, David Sherratt of the University of Oxford, said the very low response was "disappointing", and that further action must await the society's next committee meeting in mid-November. □

Researchers contest reports of tree death

Munich. Controversy has broken out in Germany following the publication of a report from an independent forestry research institute, which, its authors argue, contradicts dire warnings that forests in Europe — particularly Germany and the Czech Republic — are sickening and dying.

The report is based on 22 studies in 12 countries, which show that tree growth in Europe has actually increased over the past few decades. It has been published by the European Forest Institute (EFI), an independent research institute based in Finland, and released during EFI's annual meeting in Freiburg, Germany.

But newspaper reports highlighting statements made at the press conference have upset many other European forestry scientists and conservationists. They argue that the interpretation of the studies is invalid, as they measured only the height and diameter of trees, and did not consider other parameters relevant to the health of trees such as soil conditions and loss of leaves or needles.

The EFI receives half its funds from the government of Finland, for which forestry is a major industry, and half from 67 scientific institutes in other European countries. According to its new report, there is no clear growth trend of trees in the most northern regions of Europe. But a positive trend can be clearly seen in most regions of central Europe and in some in southern Europe, where human population is denser and pollution greater.

The report suggests that possible causes

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Pining away? The disputed study suggests increasing tree growth in central Europe.

of faster growth could be increased levels of soil nitrogen and atmospheric carbon dioxide from traffic exhaust, local climate changes, or the fact that many of the forests observed are relatively young, having been replanted following extensive tree-felling after the Second World War.

Critics of the report accept the basic data. But they argue that the rate of tree growth gives no indication of health of trees. "The trees are simply growing themselves to death," says Hubert Weinzierl, chairman of the German conservation group, BUND, who interprets the data as suggesting that

the increased growth is a reaction to their intrinsic damage.

This position is supported by the Research Centre for Forest Ecosystems at the University of Göttingen. Michael Brede-meier, its scientific secretary, argues that the increased growth of trees is associated with "chronic shock and weakness". But Heinrich Spiecker, who co-authored the EFI report, in turn questions the validity of data that have stimulated fears that German forests are dying.

The forestry department of the German federal ministry of agriculture, for example, has been monitoring the loss of leaves and needles from trees in 5,000 sites in Germany since 1983. Its latest figures indicate that more than 60 per cent of the trees under observation were damaged, as assessed by such loss, and a quarter badly damaged.

But Spiecker says that there is no proof that leaf and needle loss indicates that trees are dying. He concedes that the data in the EFI report are open to interpretation. Nevertheless, he says, he expects to see "no catastrophic loss of forest in the near future".

The European Commission disagrees. According to a recently published survey of forest conditions in the European Union, which has been running since 1987, one in five of all trees at specified sites throughout the union show clear signs of leaf or needle damage, and the extent of damage is greatest in central Europe. Thus damage appears to correlate with increased growth.

Alison Abbott

UN biodiversity advisers warned: stick to science

London. The chairman of a panel of scientific advisers to the United Nations (UN) biodiversity convention, signed at the Earth Summit in Rio de Janeiro in 1992 to provide a framework for protecting global biodiversity, has urged his members to restrict their advice to scientific issues, amid signs that some are using the body as a political forum.

Peter Schei, chairman of the convention's Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA), repeatedly urged members during a meeting in Montreal, Canada, earlier this month to avoid digressing into discussions on the politics of biodiversity. The subsidiary body was neither a "mini-conference of the parties" nor a "[policy] drafting group", said Schei, who is a biodiversity adviser to Norway's environment minister, Torbjørn Berntsen. "We are a scientific advisory body." SBSTTA, he added, needs to improve its cooperation with the scientific community.

Among other topics, the meeting considered the monitoring and assessment of biodiversity, practical approaches to taxonomy, and the economic valuation of biodiversity.

But sharp disagreement emerged during discussions on agricultural biodiversity, involving the need to balance the preservation of forests and other natural ecosystems against pressures to clear land for farming. There was also disagreement on the issue of indigenous knowledge and biodiversity conservation.

Nevertheless, Schei says the meeting was less discordant than he had envisaged, while acknowledging that government-appointed scientists often find it difficult to remain impartial in an area with a strong political dimension. For example, he says, some scientists see the question of indigenous knowledge as inevitably involving the protection of the rights of indigenous people. "That is taking the job of SBSTTA too far."

The biodiversity convention came into force in December 1993 and now counts more than 150 countries as 'parties' — those that have signed the convention and ratified it in their national parliaments. A spokesman for the International Institute for Sustainable Development, an environmentalist group based in Canada that closely

follows all UN environmental agreements, said that SBSTTA has an additional handicap: the fact that some of its members are not scientists at all, which is putting the body's scientific authority under strain.

According to some participants, delegates often arrive with national priorities in mind and are the same people who attend the annual meeting of countries that have signed the biodiversity convention, known as the Conference of the Parties. But Ann Brocklehurst, a spokeswoman for the convention, says such countries "are in a minority".

Schei agrees that scientists advising the biodiversity convention face similar difficulties to those advising the climate convention: having to advise governments on issues whose scientific underpinnings may be relatively weak. But he says he would not want to discourage the practice of governments sending the same group of delegates to both the scientific and political segments of the convention. "People who go to both have a better understanding [than raw scientists] of the transfer of scientific know-how into sound policy-making."

Ehsan Masood

Japan to double university project grants

Tokyo. Japan's Ministry of Education, Science, Sports and Culture (Monbusho) is seeking to double the budget next year of a programme that will support university research with hundreds of five-year project grants, which will each be worth an average of about US\$1 million a year. The move is part of a broad effort by science-related ministries to boost government spending on science and technology by 50 per cent over the next five years.

The programme, 'Research for the Future', was introduced this fiscal year by the Japan Society for Promotion of Science (JSPS), a semigovernmental organization run by the ministry and well-known internationally for its postdoctoral fellowships and exchange schemes for foreign scientists wishing to do research in Japan.

Less well known is its role in funding domestic postdoctoral fellowships for Japanese university researchers. In line with both responsibilities, the new programme includes support for several hundred postdoctoral 'research assistants', and will also involve participation by foreign scientists.

JSPS won an unusually large initial budget of ¥11 billion (US\$100 million) for the programme in the current fiscal year,

which began on 1 April, and, earlier this summer, review committees selected 117 projects in 17 research fields that will each receive on average about ¥100 million a year for the next five years.

Next year, in an unusual move, the ministry hopes to double this budget to ¥22 billion, as revealed in its budget request to the Ministry of Finance, submitted at the end of last month (see table, right). If accepted, as seems likely in the present political climate, JSPS will have more than trebled its total budget, from ¥12.4 billion in 1995 to ¥41.1 billion in 1997, making it a much more powerful player in Japan's university research.

The large increase is possible because the programme is classified as 'capital investment', which is usually funded by national bonds rather than taxes. As such, it is free from the ceilings normally imposed on science budgets by the Ministry of Finance. This is the first time that capital investment funding, normally reserved for public works projects, has been applied to university research grants and the payment of salaries

for young researchers, and indicates the seriousness of the government's commitment to a significant increase in funding for science.

The laws governing JSPS had to be revised to allow capital-investment funding. This legislation was passed at the end of May, and a programme committee headed by Hiroyuki Yoshikawa, president of Tokyo University, subsequently set up 17 subcommittees to select projects for universities

Highlights of 1997 science budget request for Japan's Ministry of Education, Science, Sports and Culture

	¥ billion	% change
Grants in aid of research	114.1	+12.1
Japan society for promotion of science	41.1	+56.1
(Research for the future)	(22.0*)	(+100.0)
Joint research with industry	48.3	+38.4
Academic information networks	36.5	+10.2

* New "capital investment" budget funded with national bonds
† ¥110 = US\$1

and Monbusho-affiliated institutes throughout Japan.

These programmes include projects in new materials, molecular-scale surface dynamics, next-generation process technology, computer science, synthesis science, micromechanics, biotechnology, human genome research, cell signalling, bioinformatics, brain research, structural biology, life science, developmental biology and medical engineering. Other items in Monbusho's budget request also receive large increases, although none as much as the JSPS. Competitive 'grants in aid of research' continue their rapid growth, with a proposed 12.1 per cent increase for next year to ¥114.1 billion. This combined with the new JSPS scheme will provide substantial new funds for university research.

The budget for universities to carry out joint or commissioned research with industry also receives a substantial boost of 38.3 per cent. But donations from industry are expected to remain flat because of the lingering effects of the recession.

Computer networks and new electronic information systems for libraries continue to gain large increases in their budgets, with plans to upgrade the maximum capacity of the national network linking universities to 150 megabits per second in key sections. Budgets for large facilities under construction, such as the infrared telescope in Hawaii, the B-factory in Tsukuba science city, and the helical fusion device in Gifu Prefecture also get substantial increases.

As part of a commitment to double the number of postdoctoral fellows working in Japan to 10,000 by the end of the decade, the JSPS also requests a large expansion in its fellowship programmes. This includes an increase from 420 to 750 in postdoctoral fellowships for foreign scientists to come to Japan (see *Nature* 383, 195; 1996).

David Swinbanks

Accelerator complex gains momentum

Tokyo. Monbusho's budget proposal for next year (see above) includes a relatively small request of ¥167 million (US\$1.5 million) to begin the establishment of a huge new complex for research using high-energy accelerators in Tsukuba science city, northeast of Tokyo. The complex will employ about 800 people, and will draw together researchers from the National Laboratory for High Energy Physics (KEK) in Tsukuba with Tokyo University's Institute for Nuclear Study (INS) and its Meson Science Laboratory.

Two new institutes will be formed. One will carry out research into nuclear and fundamental particle physics, and another will be for the study of the structure and function of materials. These will be built around an accelerator research facility, including the B-factory now under construction by KEK and a 50-GeV proton synchrotron proposed by INS (see *Nature* 376, 454; 1995).

The formation of the research complex comes at a time when KEK has been in need of a new purpose, following the shutdown of the TRISTAN electron-positron collider. At the same time, INS, which is based in the northwest suburbs of Tokyo, also needs rejuvenation and new facilities. One aim is to broaden the

research of the old organizations to encompass application of accelerators to materials science and structural biology.

INS and the meson laboratory will cease to be part of Tokyo University, and the institutes and complex will take on the status of national institutes for joint university use, similar to that of the complex of three such institutes for basic biology, molecular science and physiology in Okazaki near Nagoya. As such, the former Tokyo University organizations can look forward to better funding. But they may find it harder to recruit high-quality graduate students because of their loss of a direct link to Japan's leading national university.

The proposal to build a 50-GeV proton synchrotron, which is expected to cost about ¥75 billion (US\$680 million), has yet to win the go-ahead from the government. But observers say that the project is almost certain to get approval once establishment of the complex begins. Hirotaka Sugawara is widely tipped to become director general of the whole complex, while Sakue Yamada, who is currently director of INS, is expected to head the institute for nuclear and fundamental particle physics.

D. S.

Gore announces boost for neutron research at Oak Ridge lab

Washington. Al Gore, the US vice-president, brought some election-year largesse to his home state of Tennessee earlier this month when he announced that the Clinton administration is to ask Congress for \$23 million for work on a new source of research neutrons at Oak Ridge National Laboratory (ORNL) in the fiscal year 1998 (which begins on 1 October 1997).

Gore, a former Tennessee senator, also announced during a campaign stop at Oak Ridge that the administration is to provide \$10 million to upgrade the High Flux Isotope Reactor (HFIR), the most intense source of neutrons available for research use in the United States.

Although he did not specify what the money would pay for, a Department of Energy (DoE) official said the department would continue funding the addition of a 'cold' neutron capability to the reactor, extending HFIR's neutron scattering capabilities.

ORNL officials remain confident that DoE will fund the remainder of a \$70-million upgrade package they have proposed for the 30-year-old HFIR. Many of the improvements are designed to expand the reactor's capacity to produce radioisotopes.

The proposed \$1-billion new neutron spallation source planned for Oak Ridge replaces an earlier \$3-billion project to build what would have been the world's largest research reactor, which was cancelled last year in the face of tight budgets.

Bill Appleton, associate director of the laboratory, calls the announcement "a marvellous shot in the arm for the neutron community", saying that the laboratory considers it a commitment to proceed with construction of the new facility. But he acknowledges that the announced funding is considerably less than the \$43 million that ORNL considered necessary in fiscal year 1998 to proceed on the shortest possible timetable for completion.

According to Appleton, current efforts to complete a conceptual design are scheduled for completion during fiscal year 1997. At that point, assuming approval from Congress and then the president, the project would move into the detailed design phase.

Appleton argues that the US neutron research community has recognized the need for improved neutron-scattering capabilities since the 1970s, when newly built European reactors 'leapfrogged' US capabilities. Neutron scattering has "exploded" since HFIR was built in the 1960s. But Europe, with double the capability, has been able to meet the demands of university and industry researchers much more than the United States, he said.

David Kramer

Biogeochemists and industry search for common interests

London. Basic researchers in biogeochemistry from around the world last week joined representatives of industry and funding agencies to identify industrial and environmental problems that their research can help to solve.

"The reason for the meeting is the perception, common in many countries, of a change in scientific funding processes", with more emphasis on applications-oriented research, says Max Coleman, professor of sedimentology at the University of Reading, where the four-day event was held.

One of the aims of the meeting was to provide basic researchers with what Coleman describes as "the road map of success", reminding them that the new funding yardsticks of "wealth creation" and "quality of life" do not exclude basic research.

Industry was urged to outline what it wants from basic researchers. "Industry knows that basic research is its lifeblood", says Coleman but has been progressively reducing its research and development budget, which it increasingly considers the responsibility of governments. Yet governments were now more concerned with spending money on problem-solving research, such as Technology Foresight. Basic researchers had been caught in the middle, Coleman says, but would be mistaken to believe that they have lost out.

Delegates to the meeting included representatives from BP, the UK radioactive waste disposal agency NIREX, the US Department of Energy, the UK environment agency and various universities. It focused, in particular, on the potential applications of microbiology.

According to Phil Long, programme manager at the US Department of Energy

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

NIREX

Testing the ground: current priorities for funding need not rule out basic research.

Pacific Northwest National Laboratory, various conference working groups outlined a number of areas in which microbial processes could improve understanding of challenges, including underground radioactive waste disposal, oil and mineral exploration, water resources, contaminated land and even the origins of extraterrestrial life.

Frank Wobber, programme manager at the US Office of Energy Research and one of the speakers, says he sees no contradiction in the idea of setting priorities for basic research. Prioritization, he says, does not exclude the possibility of surprise results or "serendipity", which is sometimes described as the hallmark of basic research.

The idea that "it is a sin for basic research to help solve society's problems" needs to be dispelled, says Wobber. "What is wrong with a publicly funded basic researcher asking: 'is there a spin-off from what I am doing; is there a benefit?'" He also predicts that funds for "pure" basic research will be increasingly difficult to obtain without an applied dimension. "It is going to get harder and harder to say: 'I want to study bees' wings' without considering there is a huge honey industry out there." **Ehsan Masood**

East Asian research universities explore links

Tokyo. More than a dozen presidents of universities in Japan, the Chinese mainland, Taiwan, South Korea and Hong Kong met in Tokyo last week to discuss future activities of the recently formed Association of East Asian Research Universities, the first substantial attempt to link together universities in the region.

The association was launched earlier this year by Chia-Wei Woo, president of Hong Kong University of Science and Technology (HKUST). The original idea was to link together a small number of universities from East Asia with strong interests in science and technology, such as HKUST and South Korea's Pohang University of Science and Technology (POSTECH).

At the first meeting of nine presidents in Hong Kong, however, including those of

Tokyo University and Tsukuba University in Japan and Tsinghua (formerly spelt Qinghua) and Fudan universities of China, it was decided that the association could not be confined to science and technology, as these universities also have faculties in areas such as law and economics. It was therefore agreed that it should be an association of 'research' universities and so has rapidly expanded in size. Fourteen presidents met last week, and three more universities — including Kyoto University of Japan and Taiwan University — are expected to join soon.

Nevertheless, the focus is still on science and technology. Last week, two task forces were set up. One will look into areas of possible collaboration in molecular biology and biotechnology. The other will investigate computer science.

David Swinbanks

Uncertain CERN cash means UK physicists face grant freeze

London. University astronomers and physicists funded by Britain's Particle Physics and Astronomy Research Council (PPARC) have been told that no research grants will be made available to them in the next round of awards, due to start on 1 October, except in cases of "essential continuity or genuine hardship".

This decision, announced in a letter sent by the council to university research departments, reflects the uncertainty that is hanging over the funding agency about the extent to which the government will be prepared to meet the additional costs of Britain's membership of the European Laboratory for Particle Physics (CERN).

In particular, PPARC is seeking an extra 'pulse' of about £60 million (US\$90 million) over eight years, on top of its current budget, to help cover its contribution to the construction costs of the laboratory's Large Hadron Collider (LHC) (see *Nature* 381, 102; 1996). But the government's response is unlikely to be known until early next year, and research council officials have therefore decided to hold back any further research awards until the next round, in April 1997.

Although the decision will not affect the contributions that have already been agreed to major astronomy projects, both national and international, it will have a significant impact on the funding of smaller university-based projects. "This means that all the bright new ideas that we have are not going to be funded," says Malcolm Longair, Jacksonian Professor of Natural Philosophy at the University of Cambridge and president of the Royal Astronomy Society. "It's the seed-corn for the future that gets stifled on these occasions; that is where it really hurts."

PPARC officials say that, given the uncertainties over future commitments to CERN, they are choosing to exercise "a great deal of caution" over research awards. They are aware of a previous crisis that hit the research council's predecessor, the Science and Engineering Research Council, when excessive commitments — based on an expectation of extra government funds that never materialized — was partly blamed for the early closure of the Nuclear Structure Facility at Daresbury.

PPARC has also decided to hold back new four-year rolling grants to university-based particle physics groups. In the past, these have been renewed every two years; this time, researchers' salaries will be guaranteed at present only for a further two years, while funds for operating costs will be extended only for a further year, and at 70 per cent of last year's level. □

Tribal groups attack ethics of genome diversity project

Washington. Citizens' groups in the United States and representatives of tribal peoples in poor countries this week attacked proposals for the Human Genome Diversity Project, which plans to collect systematically and make available for research genetic information about people from different ethnic groups.

The critics told a panel of the National Research Council (NRC), chaired by William Schull of the University of Texas at Houston, that the proposal should be suspended on ethical grounds. "We're not interested in the patenting and commercialization of our genes, so we're against it," said Leonor Zalabata Torres of Sierra Nevada in northern Columbia.

Torres was one of the representatives of the Arhuaco people who came to Washington on Monday to testify before the panel. They alleged that scientists had already taken thousands of samples from tribal people in Columbia without meaningful informed consent, and without consulting with tribal leaders (see *Nature* 381, 11–14; 1996).

According to Abadio Green Stocel, president of the National Indigenous Peoples' Organisation of Columbia, the samples had been given on the basis that they would be used "to analyse the health of the communities". They ended up, however, in gene banks at the Centre for Disease Control and the National Institutes of Health (NIH) in the United States.

The NRC panel had been asked by the National Science Foundation (NSF) and the NIH to look at various proposals that exist to bring together research in this field under a Human Genome Diversity Project. The NSF is currently inviting grant proposals for similar research under a small pilot programme. If the NRC panel finds that a larger, more systematic programme is justified, the NSF and NIH are likely to take its advice and move to establish one.

The NRC panel will examine the case for such a project and the difficulties in setting it up. It may also estimate its likely cost and duration. Eric Fischer, head of biology at the NRC and director of this study, says ethical concerns, such as ensuring informed consent, are "a substantial aspect" of the study.

Proposals under consideration include a summary prepared by the Human Genome Organization, the international genome group, and output from scientific workshops held at Stanford University in California in 1992 and in Sardinia, Italy, in 1993.

John Moore, a Seattle businessman who was involved in lengthy and ultimately unsuccessful court action after his genes were patented by the University of California at Los Angeles, the Genetics Institute and Sandoz, the Swiss pharmaceuticals company, also testified against the proposed project, saying that gene patenting had left him "violated as a human being".

Colin Macilwain

First plant genome sequencing planned

Washington. Four US federal agencies are about to announce an international effort to decode the genetic material of the fast-growing plant *Arabidopsis thaliana* — known as 'laboratory cress' — in a move that represents the first complete sequencing of a plant genome.

The National Science Foundation (NSF) will lead the US contribution to the three-year sequencing effort, which will also involve the US Agriculture Department (USDA) and the National Institutes of Health (NIH). The project will be carried out in collaboration with Japan's Kazusa DNA Institute in Chiba Prefecture, east of Tokyo, and a consortium of 17 European laboratories funded through the European Commission.

According to officials in Washington, the NSF will provide \$9 million for the US portion of the effort, with DoE's basic energy sciences unit contributing \$2.1 million and USDA another \$1.5 million. NIH will contribute its expertise in

genome sequencing, but no funding.

Information from the sequencing is expected to be used to engineer plants to produce substances ranging from vaccines to chemicals and plastics. As well as boosting understanding of grain development, the project will lead to improvements in bioremediation.

The latter is an area of particular interest to DoE, in view of the huge clean-up tasks it faces around nuclear facilities. The energy department is also interested in the project's implications for the growing use of biomass for energy production. The project is expected to take six years to complete.

In addition to being a rapid grower, with a seed-to-seed cycle that can be as short as three weeks, *Arabidopsis* also has the smallest genome size of any plant, comprising of only 100 million base pairs, or 100 megabases. By comparison, the human genome consists of 3 billion base pairs.

David Kramer

'Mad cow' politics tries to corral science

Paris. Science was this week once again called to the aid of politicians in Europe's 'mad cow' crisis, when the British government asked the European Commission to revise a previous agreement under which Britain would reduce the incidence of bovine spongiform encephalopathy (BSE) by culling 127,000 cattle.

The UK government has been reviewing the cull following the publication of an analysis suggesting that the epidemic phase of BSE would end around 2001 without any culling, and that, given the aetiology of the disease and uncertainty about which cattle are infected, its decline would not be greatly accelerated by any culling policy short of the slaughter of around a quarter of the 10 million cattle in the national herd.

But the interpretation of this analysis, which also puts estimated numbers on the level of infected cattle remaining in British herds, has also become a centre of controversy, with others contesting the claim that it undermines demands for a broad cull.

Speaking at a meeting of European Union (EU) farm ministers on Monday (16 September), Douglas Hogg, the UK agriculture minister, said that the paper concerned showed that "BSE will in any event die out in 2000 or 2001", and that there was no culling policy that "would make a substantial difference to the rate of decline".

The cull was originally agreed to by John Major, the prime minister, at the European Union summit in Florence in June, as the price to pay for an eventual lifting of the EU ban on the export of UK beef products. But Major needs parliamentary approval before the cull can go ahead, and this looks increasingly unlikely, given growing opposition among Tory members of Parliament.

There is also uncertainty about how the opposition Labour party will vote (opposing the cull would also win votes among farmers). With the prospect of an immediate lifting of the ban seeming increasingly remote, Major seems to be gambling on satisfying domestic interests at the risk of inflaming other EU governments, by demanding that the cull be reduced or abandoned.

The government is relying on the results of a study published by Roy Anderson, professor of zoology at the University of Oxford, and both academic colleagues and scientists working for the UK Ministry of Agriculture Fisheries and Food (MAFF). They used sophisticated modelling techniques and previously confidential MAFF statistics to analyse the transmission dynamics and epidemiology of BSE (see *Nature* **382**, 787; 1996). They also compared the 'efficiency' of 14 different culling strategies, defined by the number of cases they eliminate for the total number of cattle culled.

But the European Commission disputes that the results undermine arguments in

favour of the planned cull. The paper's estimates of the "efficiency" of various culling strategies — in terms of how many cattle need to be killed to get rid of BSE cases — are of no interest to the commission, says Gerard Kiely, a spokesman for Franz Fischler, the agriculture commissioner. "Whether it is 10,000 or 2 million cattle is irrelevant to us; what is relevant is reducing BSE [cases] as fast as possible," he says.

In contrast, Kiely argues that new evidence of maternal transmission of BSE — which, as previously announced, has been submitted to *Nature* — could, in fact, require an enlargement of the cull. The

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Mulling the cull: Hogg is trying to persuade Europe that a cull will make little difference to the rate of decline of BSE.

commission is open to reviewing the culling policy to take into account such data, he says, as it would allow "taking out more cases of BSE".

Like other critics, Kiely also points out that interpretations of the Anderson paper other than that of the British government are possible, and that the evaluation of the various culling strategies could similarly justify not a reduction but an extension of the planned cull. According to the Anderson paper, the agreed cull of 127,000 cattle would eliminate 1,580 cases of BSE. But by adding a maternally targeted policy and killing 150,000 cattle, 2,380 cases would be eliminated. "I think many member states would say that is the one we need," says Kiely.

Some scientists are also concerned that the British government has selected an interpretation of the paper that suits its political needs — reinforcing the view that governments tend to accept scientific advice when it suits them and to reject it when it does not.

Mark Savey, a leading French epidemiologist on BSE, and a member of the BSE working group of the European Commission's scientific veterinary committee, also argues that the rapid interpretation given to the findings raises the wider problem of

what he claims is insufficient discussion of research results in the BSE crisis.

"Non-scientists think that when a paper is published in a prestigious journal, it is the indisputable single truth — that the matter is settled," says Savey. "But publication is only the first step; there is no paper in the BSE crisis that will close the discussion, and we will be discussing this disease for another decade."

In support of his arguments, Savey points out that much of the current literature focuses on the 'epidemic phase' of the disease. "It is a very British way of looking at things," he says. The British "say that BSE will be finished as an epidemic in 2001. What they don't say is that the disease will not have disappeared completely."

The scientific issue is to know whether the disease will remain sporadic, and how will this be handled, he says, pointing out that low levels of the disease in France has been enough to panic consumers. In a sporadic phase, BSE might evolve to find other routes of transmission, he warns, adding that the scientific community needs to "start thinking now" about these problems.

Savey also argues that, whereas the Anderson paper is "absolutely superb", the scientific aspects of the cull are merely one consideration. While scientific advice may influence culling policy, it would be simplistic to believe it should determine it, he says, arguing that this can only be done by also taking into account political, economic and other factors.

Certainly science, up to now, has had little to do with the decision to cull UK cattle. The cull was agreed by EU member states largely as a means of restoring public confidence in beef. But from the outset it has been opposed as lacking scientific justification by many scientific advisory bodies, including the British government's Spongiform Encephalopathy Advisory Committee (SEAC) and the World Organization for Animal Health.

Some British researchers argue, however, that the confirmation that any cull short of a massacre would not eradicate BSE any faster than doing nothing only lends weight to existing scientific criticism of the cull. "I suppose that is what the government is looking at, and I think the science supports it," says one of the paper's authors.

The epidemic should be allowed to die out naturally, given that precautions have been taken by abattoirs to prevent infected animals entering the food chain, says Ray Bradley, a MAFF scientist who is a member of SEAC and also chairman of the BSE working group of the European Commission's scientific veterinary committee. "The cull has no scientific basis," he says baldly. "It is only there to appease the other member states of EU."

Declan Butler

European Southern Observatory ends building row with Chile

Munich. The Chilean senate has ratified an agreement with the European Southern Observatory (ESO), based in Garching in Germany, bringing to an end a six-year dispute over the construction on Mount Paranal in northern Chile of ESO's Very Large Telescope (VLT). This had been threatened by a land dispute and arguments about the rights of Chilean astronomers to privileged viewing time (see *Nature* 374, 755; 1995).

The agreement, which supplements ESO's original agreement to develop telescope sites in Chile, confirms ESO's immunity from both local and national laws, in recognition of its status as an international organization. This means that it cannot be required to stop building work, as it has been in the past, while the local land dispute — or other similar conflicts — continues in court. In turn, in addition to ensuring privileged viewing time, Chilean scientists will benefit from agreement by ESO that it will respect national labour laws conferring collective bargaining rights. □

Malaria vaccine claims disputed

London. Studies in Thailand have found "no evidence" that the anti-malaria vaccine SPf66, developed in Colombia by the physician Manuel Patarroyo, is effective against the disease. Earlier trials in Colombia and Tanzania had indicated an effectiveness of 34 and 31 per cent respectively. But a larger study carried out in Thailand by Thai, British and US researchers, and involving 1,221 children between the ages of 2 and 15 living in a remote refugee camp, divided into two groups, has now found that roughly the same number of children in each group developed malaria, and that the severity of the infections was also broadly similar. According to the researchers, writing in *The Lancet* (348, 701; 1996) last week, the results, which corroborate a smaller trial in the Gambia last year, suggest that "there appears to be little justification for further trials with this vaccine". □

Solar summit attracts few stars

Paris. The first UN world summit on solar energy opened in Harare, the capital of Zimbabwe, on Monday (16 September). The meeting, which is being organized by Unesco, hopes to create and find funding for a World Solar Programme 1996–2005, which would support renewable energy projects designed to provide cheap energy in remote rural areas. About 2.4 billion people live without access to electricity. But only nine heads of state turned up for the summit — hosted by Robert Mugabe, president of Zimbabwe — five of them from Africa, while many others sent only energy or science ministers, or local diplomats. □

Drug companies plan venture fund

Basel. The Switzerland-based pharmaceutical companies Ciba-Geigy and Sandoz, which are to merge into a single company called Novartis later this year with a reduction of 10,000 jobs world-wide, are setting aside SFr100 million (US\$82 million) as start-up venture capital for any scientists wishing to turn entrepreneur. François L'Éplattenier, former head of research at Ciba-Geigy and a member of the Novartis board of directors, will administer the funds. Further details of who will be eligible, and what type of venture will be supported, will be announced next week. □

Skull of 'Xhosa king' is European

Cape Town. A self-styled South African chief, Nicholas Gcaleka, has been charged with fraud for failing to pay grocery and other bills. The charges came shortly after the completion of tests revealing that a skull claimed by Gcaleka to be that of the

late Xhosa King Hintsa in fact belonged to a European woman.

Gcaleka travelled to Scotland earlier this year and returned with a skull that he claimed was that of King Hintsa. The king was shot at point-blank range on the orders of Colonel (later Sir) Harry Smith, while pleading for mercy during the frontier war in the Eastern Cape Province in 1835. But on Gcaleka's return to South Africa, the Xhosa royal house confiscated the skull, and handed it over to scientists for forensic tests.

Independent morphometric analyses by Phillip Tobias, professor emeritus of anatomy at the University of Witwatersrand, and Vincent Phillips, professor of oral pathology at the University of Stellenbosch, indicate that the skull is that of a female of Caucasoid descent. The skull is to be returned to Scotland. □

Israeli woman wins embryo case

Jerusalem. A woman whose husband left her after his sperm was used to fertilize her eggs *in vitro* can try to have the frozen embryos brought to term in a surrogate mother, Israel's Supreme Court ruled last week. The 7:4 decision, with the majority basing its arguments largely on ethical rather than legal grounds, ended a five-year legal battle between Dani Nahmani, who had argued that he could not be required to father a baby by a woman from who he is now estranged, and Ruti Nahmani, who claims that the frozen embryos represent her only chance of becoming a mother. □

Rabbit virus set for release

Sydney. The lethal rabbit calicivirus (RCV) disease is to be released in Australia in a nationwide campaign to reduce the annual A\$600-million (US\$472-million) damage that rabbits cause to agriculture. A release at 280 sites has been announced by John Anderson, the federal minister for primary industries, after a year-long controversy following the accidental spread of the disease from test sites on a supposedly isolated island off the mainland in South Australia.

Scientists are still uncertain how the disease has been able to spread so fast. The Division of Wildlife and Ecology of the Commonwealth Scientific and Industrial Research Organization (CSIRO) conducted the field tests, which were welcomed by farmers but opposed by some groups that cited warnings by two US researchers that the virus could jump the species barrier. No evidence for this has been found in laboratory tests with both native and introduced species. □

Altai objects to spacecraft debris

Moscow. Russia's Security Council, headed by General Aleksandr Lebed, has received a letter from the Altai Republic State Ecological Committee asking it to evaluate the potential dangers of rockets launched from the neighbouring 'cosmodrome' Baikonur (now in Kazakhstan), which either fall back to Earth or burn up in the atmosphere over Altai territory. According to a recent study, 1,250 tonnes of metal parts have fallen on the republic's two regions, which have a combined area of 7,400 square kilometres. This airborne 'garbage' is toxic because of the rocket fuel it contains, and therefore presents a threat to ecological security of the region, according to local officials. □

Less cash from French AIDS event

Paris. French AIDS associations and researchers have been dismayed to learn that the amount of money raised by this year's Sidaction — a fund-raising event involving all national television channels — amounted to just FF65 million (US\$13 million), compared to FF300 million in 1994. Some of the blame is being attributed to militants from the pressure group Act Up, who were responsible for aggressive outbursts during the televised show. Others attribute the event's poor returns to its emphasis on the need for donations rather than attempting to inform viewers about the AIDS issue, as well as a general decline in interest in AIDS among the public in France. □

Spending on BSE research

SIR — As one of the industry-independent representatives on the Government's Priorities Board for Research and Development (R&D) in Food and Agriculture from 1990 to 1994, I chaired the Ruminants Advisory Sectoral Group (RASG). The group advised the government through the priorities board on R&D in the various areas in the ruminant sector, which included animal welfare and disease, and had access to information on government and other research projects and their associated funding.

It is in that previous research role that I am responding to the article on bovine spongiform encephalopathy (BSE) (*Nature* **382**, 755–756; 1996) by David Skegg. His article accompanied the publication of the work of Anderson *et al.* (**382**, 779–788; 1996). Although Skegg summarizes the situation as accurately as is possible in the light of the long-term nature and complexities of the disease and associated R&D, the statement in his final paragraph about spending on BSE research is inaccurate and misleading.

Skegg states that “during the three years 1988 to 1991 the UK government invested the paltry sum of £3 million on BSE-related research”. There have always been several sources of funding for government BSE research, and it would appear that Skegg has not had access to all the information relating to BSE R&D and its funding for 1988 to 1991. The information collated by the RASG in 1991 indicated annual expenditure from all sources of considerably more than £3 million. The RASG report for 1991–92 shows that, of the total spent on the ruminant welfare and disease area of £20.134 million, £8.324 million was for BSE and related scrapie work.

It is interesting that the BSE R&D expenditure was some 20% of the total covered by the ruminant group, which included work on forages, nutrition, reproduction and genetics, meat and milk production, and production systems, as well as other disease and animal welfare issues.

The Ministry of Agriculture, Fisheries and Food (MAFF) document *Research Strategy 1996–2000* shows its expenditure on BSE research for 1996–97 to be £10.400 million. That is some 66% of the total MAFF expenditure of £15.858 million in the animal health and welfare sector. This is complemented by further R&D by other sponsors such as the Biotechnology and Biological Sciences Research Council, the Medical Research Council and the Wellcome Trust.

Ian Howie

Ryfield,
Wormbridge,
Hereford HR2 9DB, UK

SIR — The British and European Union (EU) authorities were recently accused by

D. W. van Bakkum and P. J. Heidt (*Nature* **382**, 574; 1996) of having missed an early opportunity of resolving the issue of the risk to humans from bovine spongiform encephalopathy (BSE) by their alleged failure to initiate and fund challenge experiments in chimpanzees available in Europe.

There were, however, discussions involving UK and EU scientists and authorities in 1989 and 1990 about the feasibility of treating chimpanzees with material from BSE-infected cows. As a virologist and deputy director of the Dutch primate centre at that time, I was directly involved in some of those discussions.

It was impossible to design conclusive study protocols answering the questions of human infectivity of the BSE agent and of the risk for the consumer of beef products, which are different issues. To show infectivity, the intracranial injection of infected brain extracts was considered. To evaluate the risk to consumers, the only relevant route is oral challenge with beef from infected animals. From epidemiological studies, it is clear that if BSE is causing disease in humans, the transmission rate is very low, needing hundreds of chimpanzees to get significant results, more than the number available in Europe.

But even if oral transmission had resulted in disease in chimpanzees, the relevance of these results for humans could be argued. It is doubtful whether chimpanzees and humans have similar sensitivities to pathogens transmitted by meat. Chimpanzees are almost exclusively vegetarian, and resistance in humans to the BSE agent may have been selected for by hundreds of thousands of years of meat consumption. At the other extreme, if animals injected with heavily contaminated material in their brains remained healthy, the relevance for humans could again have been challenged.

Although van Bakkum and Heidt state otherwise, there is no evidence that chimpanzees are the experimental animal of choice to attack this problem. Knowledge of the pathogenesis of natural infections in chimpanzees is limited, and there is experience only with a few human pathogens in this species. The most widely studied have been the hepatitis viruses and human immunodeficiency virus (HIV). And although chimpanzees have been instrumental in testing vaccines for hepatitis-B virus, the course of acute hepatitis and especially the long-term sequelae of chronic infections, which are the most relevant for the human disease, are mild or absent in chimpanzees. With two exceptions, the hundred or so chimpanzees infected with HIV failed to develop symptoms similar to human AIDS. So any negative results with BSE in chimpanzees would be difficult to extrapolate to

safety of beef from BSE-infected animals for human consumption.

In summary, the British and EU authorities considered chimpanzee experiments at an early stage of the BSE epidemic and have consulted the relevant specialists in the field. Both DGXII of the European Commission (the science directorate) and the UK government have from the beginning of the epidemic funded research into the BSE problem, and both were willing to provide the vast amounts of money that the chimpanzee experiments would have cost.

The many unanswered questions about BSE (and prions in general) have less to do with the lack of funding or the handling of the problem by the British or EU authorities than with the fact that BSE is an extremely difficult scientific nut to crack.

H. Schellekens

Diagnostic Center SSDZ,
PO Box 5010,
2600 GA Delft, The Netherlands

SIR — Van Bakkum and Heidt confirm the truth that every cloud has a silver lining — in the case of the BSE scare, an opportunity to “dispose of” numbers of chimpanzees surplus to requirements of the AIDS research programme in Europe. You recently reported a similar chimpanzee glut at the US National Institutes of Health (*Nature* **381**, 722; 1996) but in that case the proposed solution (retirement homes) seemed rather more humane than van Bakkum and Heidt's suggestion of “intra-cerebral injection of BSE brain and beef and administration of infected cow brain and meat and perhaps even milk by the cutaneous, intravenous and... oral route”.

They are of course right to say that the BSE saga is full of unknowns; neither the relation of the ‘new variant’ of Creutzfeldt-Jakob disease to BSE, the relation of BSE to animal feed (or insecticide?) or the extent of risk to the human population is established. However, a couple of other things are well known but, strangely, not yet acted on. The first is that BSE is a direct result of intensive farming methods (‘factory farming’ of animals) which are generally objectionable on animal-welfare grounds. The second is that these methods have reduced the price of meat and dairy products to the extent that we in the rich countries now damage our health by eating too much of them and damage the environment to feed our habit. So perhaps the lessons of BSE, if read aright, will lead to improvements in both human and animal welfare — another silver lining unnoticed by van Bakkum and Heidt — even if they come too late for the hapless chimpanzees?

J. Z. Turner

Crystallography Department,
Birkbeck College,
London WC1E 7HX, UK
e-mail: j.turner@mail.cryst.bbk.ac.uk

Unravelling the Indian rope-trick

SIR — Accounts of miracles appear in many religious texts, and the evidential nature of such testimony has long been debated. Where the historicity of such accounts is considered theologically important, scholars have been particularly concerned with their reliability. Sceptics have suggested that testimony became exaggerated over time¹, but the nature and extent of the evidence has not allowed this hypothesis to be tested empirically. However, while researching perhaps the world's best-known secular miracle — the Indian rope-trick — we discovered a unique set of eyewitness accounts that could be used to evaluate this hypothesis.

Descriptions of the Indian rope-trick can be found in Buddhist mythology and Hindu philosophy, and in eye-witness accounts from the fourteenth century onwards². In the classical version of the trick, which normally takes place in the open, the magician throws one end of a rope into the air, and the rope remains rigid. A boy climbs up the rope and, on reaching the top, disappears. The magician then orders the boy to return and, when he refuses to do so, the magician climbs up the rope carrying a knife and also disappears. Next, the boy's dismembered body parts fall to the ground. The magician then descends the rope, covers the body

parts, and the boy is magically restored.

Although many reports could be dismissed as travellers' tales, others cannot be dismissed so easily. In the late nineteenth century, the trick was frequently discussed in the Anglo-Indian and British press, and many individuals wrote claiming to have witnessed the trick (or a version of it) in India. These accounts prompted many people to search extensively for the trick, often offering large sums of money to anybody able to perform it. The searches proved fruitless

and the offers remained unclaimed³.

Many writers have since attempted to explain why so many people claimed to have seen a trick that no magician seemed able to perform. Some suggested that the witnesses were hallucinating, hypnotized or hoaxers. Others argued that witnesses had seen a mundane performance of Indian street magic and, echoing the sceptical argument on alleged miracles, had exaggerated their accounts over time. Despite generating a huge amount of literature, including an article in *Nature*⁴, the mystery was never resolved.

We hypothesized that if eyewitnesses

CATEGORIES OF FIRST-HAND OBSERVATIONS OF ROPE-TRICK

Category	No. of cases in category	Mean (and s.d.) in years of 'lapse'	Brief description of type of testimony in each category
1	3	4 (1.73)	Boy climbs up rope, climbs down again.
2	3	12.67 (7.50)	Boy climbs up rope, seems to vanish, before reappearing at top of rope.
3	1	31	Boy climbs up rope, vanishes completely at top.
4	10	32.50 (9.3)	Boy climbs up rope, vanishes at top, then reappears somewhere audience has not been watching (e.g., behind the crowd).
5	4	41.75 (10.53)	Boy climbs up rope, vanishes at top. Boy then reappears in a basket which has been in full view of audience.

A brief description of each category, the number of cases in each category and the mean of the amount of time between the witness seeing the trick and reporting it (measured in years and referred to as 'lapse').

New **Swedish** solutions for purifying peptides
of **any source** using **any technique**



INTRODUCING:

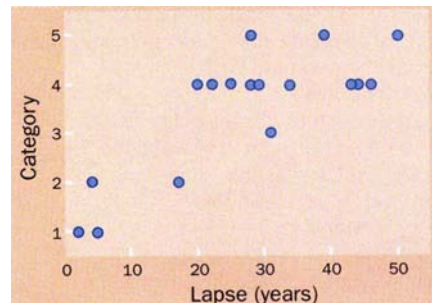
- ten new reversed phase chromatography columns
- two new ion exchange chromatography columns
- a new size exclusion chromatography column
- one completely new system for peptide, oligonucleotides and other biomolecules (ÅKTA is the Swedish word for real, it's pronounced eckta).

were exaggerating their accounts over time, one might expect a positive correlation between the impressiveness of the account and the length of time between performance and report. To test this notion, we searched all the publications discussing the rope-trick, and located 48 eyewitness accounts. Reports were rejected from our analysis if they contained second-hand testimony ($N = 4$), did not record the year the performance took place ($N = 13$), did not contain any description of the performance ($N = 6$) or contained too few details to reconstruct the alleged events of the performance ($N = 5$). There remained 21 accounts that were first-hand, sufficiently detailed and included necessary dates. These fell into five categories, which were then ordered in terms of how impressive the trick appeared to be. The order was decided independently by a panel of magicians. The table contains a brief description of each category, the number of reports in each category, and the mean of the amount of time between the witness seeing the trick and reporting it.

A Spearman's Rank Correlation Coefficient between 'lapse' (2 to 50 years) and 'category' (1 to 5) was large ($N=21$, Rho [corrected for ties] = 0.78; $z = 3.50$) and highly statistically significant (P [2 tailed] = 0.0005). This correlation is illustrated in the figure. In addition, a stepwise regression indicated that the 'lapse' variable was the best predictor of 'category' ($r = .866$) and

that this predictive power was not increased by including the year in which the trick was observed or the year in which it was recalled. In short, this analysis strongly suggests that witnesses' testimony had become significantly more elaborate over time.

This analysis casts severe doubt on the validity of accounts describing more impres-



Scatterplot of the 'category' and 'lapse' data for each of the 21 eyewitnesses.

sive performances. But how accurate are those less impressive accounts in category 1? Evidence from the account, in 1919, with the shortest lapse (2 years) suggests that even these accounts might be unreliable. One eyewitness claimed not only to have seen the rope trick but to have taken a photograph of it⁵. An examination of the photograph clearly showed a common pole-balancing trick, in which a child balances on the end of a long (10–12 foot) bamboo pole held by the performer. When challenged by a sceptic, the

witness apparently admitted this to be the case⁶. In short, the data suggest that witnesses saw an unimpressive trick (possibly the pole-balancing trick) and then added elements drawn from Indian street magic and/or mythical versions of the trick as time passed.

In 1936, the *Nature* article⁴ dismissed the testimony relating to the Indian rope-trick as the product of unreliable evidence. The author, Colonel Elliot, urged readers "to look always for a natural explanation of any phenomenon, and when one is not forthcoming, to await the advent of more knowledge, confident that a normal and not supernatural explanation is always forthcoming, provided we have the requisite knowledge". Sixty years on, a fresh approach to the evidence supports a sceptical view of the Indian rope-trick and casts doubt on testimony relating to other alleged extraordinary events.

Richard Wiseman

Peter Lamont

*Psychology Department,
University of Hertfordshire,
College Lane,
Hatfield, Herts AL10 9AB, UK*

1. Kurtz, P. *The Transcendental Temptation*. (Prometheus Books, New York, 1991).
2. Eliade, M. *The Two And The One*. (Harvill, London, 1965).
3. Siegel, L. *Net of Magic*. (Univ. Chicago Press, 1991).
4. Elliot, R. H. *Nature* **138**, 425–427 (1936).
5. Holmes, F. W. *Strand Magazine*, April, 310–311 (1919).
6. Elliot, R. H. *The Myth of the Mystic East* (Blackwood, London, 1934).

Are you working with natural peptides? Synthetic peptides? Recombinant peptides? Peptide fragments? Or all of them? Whatever peptide you work with, your options for purifying them just increased.

With ten new columns for reversed phase chromatography, you're nearly certain of finding the selectivity you need in our extensive range. All of these new RPC columns deliver high resolution; combined, they'll take you from purification and analysis to peptide mapping.

We can support you with advice and solutions for other peptide purification techniques as well. Are you separating peptides with poor solubility? Our new size exclusion column withstands high pH and solvents. Do you need an extra technique to help you with a difficult-to-separate peptide? We have two new ion exchange columns that permit very high resolution and withstand high pH ranges.

What's more, all 13 columns are supported by ÄKTA™purifier—a revolutionary new purification system for peptides, oligonucleotides and other biomolecules. Its preset protocols let you resolve all major purification tasks quickly and easily. Its control system lets you instantly transfer your methods to purification systems at all scales.

Want to know more about our peptide purification solutions? Call us: 1 (800) 526 3593 from the USA; +81 (0)3 3492 6949 from Japan; or +46 (0)18 16 50 11 from Europe and the rest of the world; or meet us on the Internet at <http://www.biotech.pharmacia.se>.



**Pharmacia
Biotech**

Uppsala, Sweden. (And the rest of the world)



IPCC report, chapter and verse

SIR — Singer's letter¹ on the Intergovernmental Panel on Climate Change (IPCC) contains factual errors and makes selective use of relevant information. This response is to correct his errors and provide a balance to his selective use of material.

(1) Contrary to Singer's claim, changes to the main IPCC Second Assessment Report (SAR)² were not made to ensure that it conformed to the Summary for Policy Makers (SPM). It is the full report that forms the basis for the summary, not vice versa.

(2) Singer falsely implies that satellite-based temperature data are not considered in the SAR. These data are not covered in the report's very brief (5-page) SPM, but they are discussed in the Technical Summary (which is meant to be read in conjunction with the SPM), and, in considerable detail, in the body of the report. That the SPM does not specifically discuss these data is merely a reflection of the unanimous judgement of the countries involved in the process that they are of insufficient importance to the policy process to be included in such a brief overview.

(3) The IPCC report, contrary to Singer's claim, does not misuse the work of respected scientists. The same caution that characterizes the papers Singer cites^{3,4} and all other publications on the subject is echoed in the IPCC report. Given that the report was written by (among many others) the authors of the cited papers, to claim that these scientists misused their own work is ludicrous.

(4) Singer cites his own letter in *Science*⁵ to support his claim that the IPCC summary is selective, and that this selectivity is based on political motives. No such selectivity exists, and the points made in Singer's *Science* letter have all been refuted⁶.

Tom M. L. Wigley

(Co-lead author for Chapters 2, 6 and 8 of the IPCC report)
National Center for
Atmospheric Research,
PO Box 3000,
Boulder, Colorado 80307-3000, USA
e-mail: wigley@ncar.ucar.edu

1. Singer, S. F. *Nature* **382**, 392 (1996).

2. *Climate Change 1995*. Vol. 1. *The Science of Climate Change*, IPCC Second Assessment Report (eds Houghton, J. T. et al.) (Cambridge Univ. Press, 1996).

3. Mitchell, J. F. B. et al. *Nature* **341**, 132–134 (1995).

4. Santer, B. D. et al. *Clim. Dyn.* **12**, 77–100 (1995).

5. Singer, S. F. *Science* **271**, 581–582 (1996).

6. Wigley, T. M. L. *Science* **271**, 1481–1482 (1996).

SIR — Hammitt *et al.* (*Nature* **381**, 301–303; 1996) incorporate a discount rate into their proposed index for assessing environmental effects of greenhouse-gas emissions. At a 3% discount rate, the proposed measure would indicate that it is good to delay the extinction of one species for 11 years even if that causes the extinction of five additional species a century hence, and that it is worth saving one additional species today

even if that causes the extinction of 10 species 78 years from now.

A non-zero discount rate may be inappropriate for long-term climate change problems involving potential reductions in biodiversity. For such problems, preservation of irreplaceable environmental assets should be paramount.

Ken Caldeira

Climate System Modeling Group,
Lawrence Livermore National Laboratory,
7000 East Ave, L-256,
Livermore, California 94550, USA
e-mail: kenc@llnl.gov

SIR — In your leading article "Climate debate must not overheat" (*Nature* **381**, 539; 1996) you speak of "continuing dissent from a dwindling band of sceptics".

Around 1980, when I first came to the conclusion that the existing climate models were overestimating by at least two- or threefold the global warming from a doubling of carbon dioxide, I essentially stood alone on this issue. Since then I have been joined by the following individuals and organizations that I am aware of: Sherwood Idso (Western Fuels Association Inc.); Jerome Namais (Marshall Institute); Robert C. Balling Jr (Cato Institute); Patrick J. Michaels (eco); Richard S. Lindzen (Heartland Institute); S. Fred Singer (Global Climate Coalition); Gregg Easterbrook (Doctors for Disaster Preparedness); and William M. Gray.

Even Robert M. White, Peter Toynbee, Roger Revelle, Chauncey Starr, C. C. Wallen and Freeman J. Dyson have at least proposed a wait-and-see attitude. At the same time, I am unaware of anyone who has publicly expressed a dissenting view and later redeclared support for the IPCC 'consensus'.

To me, this doesn't represent a "dwindling band of sceptics".

Hugh W. Eilsaesser

Lawrence Livermore National Laboratory,
PO Box 808, L-264,
Livermore, California 94550, USA
e-mail: hughel@targa.llnl.gov

Statistical leaps

SIR — The Briefing articles about athletic records (*Nature* **382**, 12–16; 1996) address, in part, the relative athletic abilities of men and women. Sir Winston Churchill's warning about the use and abuse of statistics should be borne in mind.

Use of world record lists can lead to misleading conclusions because of the performance(s) of a single outstanding individual and/or the use of enhancing drugs. It is not always possible to distinguish between these two alternatives. The graph shown for 1,500-metre world records (page 15) illus-

trates this. The men's record shows steady improvement over the past 40 years, whereas the women's record flattens out after 1980 because of an outstanding performance that year. How does one extrapolate these graphs to see where they may intersect? Plots of records for other women's events (100 metres to 800 metres) would show similar features.

Over the past decade, few women have run faster than the record set in 1976, and in this decade no woman has come remotely close to the 100-metre and 200-metre records set in 1988. Outstanding performers can skew world lists in men's events also; over the past 60 years, the 800-metre record has been dominated by three people (Harbig, Snell and Coe), and the long jump by two (Owens and Beamon).

The danger of allowing any one performance to dominate one's thoughts is best illustrated by Jesse Owens in 1935. On 25 May that year, he broke or equalled six world records at Ann Arbor, Michigan; six weeks later, he failed to win a single national crown, being beaten by Eulace Peacock in the 100 metres and long jump, and by Ralph Metcalfe in the 200 metres.

To help to identify events in which women may outstrip men in the near future, (linear) extrapolation of segments of world records to seek intersection (*Nature* **355**, 25; 1992) is fraught with problems. For the men's 1,500 metres, plots of the 'open' record and that for the over-40s for the past 20 years, followed by (linear) extrapolation, would probably lead to the conclusion the old-timers would be faster than their younger brethren some time within the next few decades. Listing the twenty-fifth best performance (say) in each event would help to eliminate 'outliers' caused by an outstanding performance and/or the use of performance-enhancing drugs. We may then obtain supporting evidence for the suggestion of researchers (*Nature* **382**, 16; 1996) as to the specific events in which women are likely to surpass men in the decades to come.

Ajoy K. Baksi

Department of Geology & Geophysics,
Louisiana State University,
Baton Rouge,
Louisiana 70803, USA
e-mail: ajoy@aapg.geol.lsu.edu

Le mot juste

SIR — The correct technical term for the genre of science fiction exemplified by the film *Independence Day*, recently reviewed by Henry Gee (*Nature* **382**, 681; 1996), is the spoonerism "thud and blunder".

Jamie Thomerson

Department of Biology,
Southern Illinois University,
Edwardsville,
Illinois 62026-1651 USA
e-mail: sthomas@siue.edu

The holes are defined by the string

Edward Witten

Must general relativity finally bow to quantum mechanics? Calculations that describe black-hole properties using collections of superstrings have gone some way towards resolving one of the most vexing puzzles in physics.

ACCORDING to Einstein's theory of general relativity, if a sufficiently large mass collapses into a sufficiently small space, a black hole is formed. To the extent that quantum-mechanical effects can be neglected, a black hole has the property that nothing, not even a light ray, can escape from it. That explains the name: an isolated black hole looks literally like a black hole in space, from which no light emerges.

Although classical black holes do not emit anything, they absorb anything that comes near enough. When an object falls into a black hole, its entire rest energy — Einstein's $E=mc^2$ — can potentially be released in the form of radiation, which can escape and be seen by a distant observer. Therefore, although an isolated classical black hole is altogether dark, if a black hole is embedded in a suitable astrophysical environment, the region just outside the hole can shine brilliantly. This has

made it possible for astrophysicists to identify candidate black holes both in our Galaxy and in distant quasars.

Black holes are also a fascinating test case for trying to combine quantum mechanics and general relativity. Physicists have found it vexingly difficult to unify, or even reconcile, these two theories. The problems are hard to explain in non-technical terms, but they are definitely there: the delicate constructions that are required in quantum field theory (the most fully known expression of quantum mechanics) seem incompatible with the requirements of general relativity. The contradiction between quantum mechanics and general relativity — which together are the basis for what we know of the fundamental laws of nature — is generally seen as a central challenge in physics, to say the least. Physicists have sought to meet this challenge through the new framework of string theory, about which we will say more later.

When one tries to think about black holes quantum mechanically, one runs into many difficulties. At the most basic level, quantum-mechanical unitarity (the requirement that the probabilities of all possible outcomes add up to one) does not permit the existence of an object that can absorb matter but cannot emit matter. This particular question was addressed 20 years ago by Stephen Hawking, who showed that, quantum mechanically, an isolated black hole is not completely black, but emits radiation at a rate which is proportional to Planck's constant $\hbar$ (and so the rate is zero in the absence of quantum mechanics)¹.

Following earlier heuristic work by Jacob Bekenstein², Hawking's discovery led to the idea that a black hole has a certain temperature and entropy at the quantum-mechanical level (generally, entropy is a measure of disorder; in standard quantum systems, it is the logarithm of

Circling the inverse square

ONE bit of physics that almost everyone knows is that the force between two electric charges, or two masses, follows an inverse square law. That is, if they are separated by a distance r then the force between them is proportional to r^{-2} .

If one takes the inverse square law seriously, one must face the fact that the force becomes infinite as r goes to zero. Early in this century, physicists realized that a classical electron orbiting an atomic nucleus should emit electromagnetic radiation and spiral into the nucleus in a finite time, driven by the singularity of the r^{-2} force at small r .

From grappling with this contradiction, quantum mechanics was born. By the mid-1920s, the Schrödinger equation gave a sensible description of atoms at the non-relativistic level. The quantum uncertainty principle effectively smeared out the electron and prevented it from seeing the singularity.

Extending this success to take account of special relativity was more difficult. It was necessary to develop quantum field theory, quantum electrodynamics and the renormalization theory (a systematic procedure for subtracting the infinities

created by the inverse square law) of Feynman, Schwinger, Tomonaga and Dyson. Finally, by about 1950, a satisfactory account of atoms could be made, including the effects of both quantum mechanics and special relativity.

A natural hope was that having understood how to treat the r^{-2} force between electric charges, one could treat the r^{-2} gravitational force similarly. That hope was frustrated: because of the nonlinear mathematics used in general relativity, the new concepts that enabled physicists to cope with electromagnetism fail dismally for gravity.

And so we face a contradiction between quantum field theory and general relativity similar to the contradictions that led to quantum mechanics. Many physicists believe that this contradiction contains the seeds of an upheaval as profound in its own way as the discovery of quantum mechanics or relativity.

String theory is the one concrete proposal for a new framework in which the r^{-2} force of quantum gravity is tamed. In string theory, elementary particles are re-interpreted as small loops of vibrating string. Their 'stringiness' gives a new kind of uncertainty, which plays a

role analogous to that of the quantum uncertainty principle, and avoids the r^{-2} singularity. As a result, gravity becomes compatible with quantum mechanics. In fact, when physicists began, more than 25 years ago, to develop string theory, they unearthed a remarkably rich and subtle structure and learned to their amazement that gravity is *required* for the consistency of string theory. So one may fairly say that pre-string quantum theory makes gravity impossible, whereas string theory makes it necessary.

Even after 25 years, the understanding of string theory is in its infancy, and surprises are the norm. It was believed for many years that there were five possible string theories, prompting the question: if one of these theories describes our Universe, who lives in the other four worlds? But recently it has become clear that those five string theories are limiting cases of one majestic and still-mysterious theory. This theory, which is sometimes called M-theory (according to taste, M stands for magic, mystery, marvel or membrane) is seen by many as a likely candidate for a complete description of nature. E. W.

the number of states). For a typical astrophysical black hole, the Bekenstein–Hawking temperature is unobservably small, whereas the Bekenstein–Hawking entropy is huge — much bigger than the entropy of an ordinary star, for instance — and difficult to interpret.

While Hawking's result perhaps eliminated the most naive paradox posed by quantum-mechanical black holes, many apparent contradictions remain. Speculations about how the puzzles will ultimately be resolved have ranged all over the map. At the risk of some oversimplification, two main competing points of view are as follows:

(1) Quantum mechanics as we know it does not work for black holes. Some modification of quantum mechanics will be necessary to understand them.

(2) Black holes will ultimately turn out to obey the standard laws of quantum mechanics. The Bekenstein–Hawking entropy of a black hole — like the entropy of any other quantum-mechanical system — will turn out to equal the logarithm of the number of quantum states of the black hole.

To make progress on these questions, one needs a theory that consistently combines quantum mechanics and general relativity, because the whole issue concerns the application of quantum mechanics to a general-relativistic object, the black hole. The only real candidate for reconciling quantum mechanics and general relativity is string theory, which has been intensively developed in recent decades with the hope of obtaining a more unified understanding of natural law. In this framework, the fundamental objects are tiny strings, which do not produce the awkward infinities thrown up by the point-like particles of conventional physics. Surprisingly, when particles are replaced by strings, gravity is an inevitable consequence.

In practice, though, the development of string theory has shed very little light on black holes until the past few months. String theory is very imperfectly understood, and until recently it has only been possible to calculate quantum gravity effects when those effects are small, which is not the case in the context of black holes.

But since the spring of last year, in an upheaval sometimes called 'the second superstring revolution', physicists have

begun to learn what happens in string theory when the quantum gravity effects are big. It turns out that 'dualities' — mysterious symmetries that generalize the relationship between electricity and magnetism to other forces — play an important role here. Perhaps the biggest resulting surprise has been that, as we now understand it, there is only one string theory. The five or six different theories that have been developed and studied independently are — it is now clear — all equivalent. They are different formulations of the same, still rather mysterious theory, each formulation being most useful in its own regime.

Lately, the new techniques have been applied to black holes, shedding at least a bit of light on the old mysteries. In a paper published in June, Andrew Strominger and Cumrun Vafa³ succeeded in counting the number of quantum states of certain black holes (carrying appropriate electric and magnetic charges) using string theory. String theory is crucial because without a microscopic theory of quantum gravity, one would have no idea where to begin in order to count the quantum states of a black hole. Technically, the Strominger–Vafa computation depended on the fact that, without changing the number of quantum states, a black hole can be deformed into a collection of 'D-branes', exotic objects that we now understand have an important role in string-theory dualities⁴. Counting of D-brane states is a well-honed art, and the relation of black holes to D-branes made the

counting of black-hole states possible.

The main result of the Strominger–Vafa computation was that the logarithm of the number of quantum states did coincide with the Bekenstein–Hawking entropy — as it should, if conventional quantum mechanics applies to black holes. The Strominger–Vafa computation was soon followed by a variety of others (refs 5–8, for example) in which the Bekenstein–Hawking entropy was compared with the number of quantum states for black holes with different angular momentum or charge, or in different dimensions of space-time. In many of these instances the classical black-hole solution was not known before the string theory computation was performed, but ultimately one finds agreement between the Bekenstein–Hawking entropy and the counting of quantum states.

These computations have given striking support to the view that black holes obey the standard general principles of quantum mechanics. But quantum black holes remain extraordinary and mysterious in many ways. Ambitious proposals concerning a new kind of black-hole 'complementarity'⁹ (echoing the more familiar complementarity of quantum and classical physics) and a holographic interpretation of the Universe^{10,11} give hints of how little we understand, even now. □

Edward Witten is in the Institute for Advanced Study, School of Natural Sciences, Olden Lane, Princeton, New Jersey 08540, USA.

NEUROPSYCHOLOGY

Pictures, words and the brain

Alfonso Caramazza

OVER the past few years cognitive neuroscientists have made great strides in charting the functional organization of the human brain. These developments have been made possible by increasingly sophisticated functional neuroimaging techniques, such as positron-emission tomography and functional magnetic-resonance imaging, which provide an index of neural activity in different areas of the brain during the performance of a cognitive task. Several notable contributions have been made by scientists at the National Institute of Neurology, London, who have concentrated largely on the neural representation of language processes. On page 254 of this issue¹ they describe their latest result, and it is an important one — a map of the areas of the brain that are involved in processing the meaning of words.

Until recently, virtually all that was known about the neural basis of language came from the study of neurologically

impaired patients. By investigating the disorders in language and cognition that are associated with specific forms of brain damage, neurologists and neuropsychologists have been able to chart the functional organization of the human brain. For example, it has been shown that damage to specific regions of the left frontal lobe frequently results in a deficit restricted to processing the syntactic structure of sentences²; and that damage to the medial and inferior parts of the left temporal lobe is often associated with a deficit in retrieving words but not in processing syntactic structure³. These and other such aphasic disorders have been used to chart the functional organization of syntactic and lexical processes in the brain.

The performance of neurologically impaired patients has also been used to inform theories of the representation of word meaning in the brain. Study of a disorder known as optic aphasia, which was first described near the end of the last

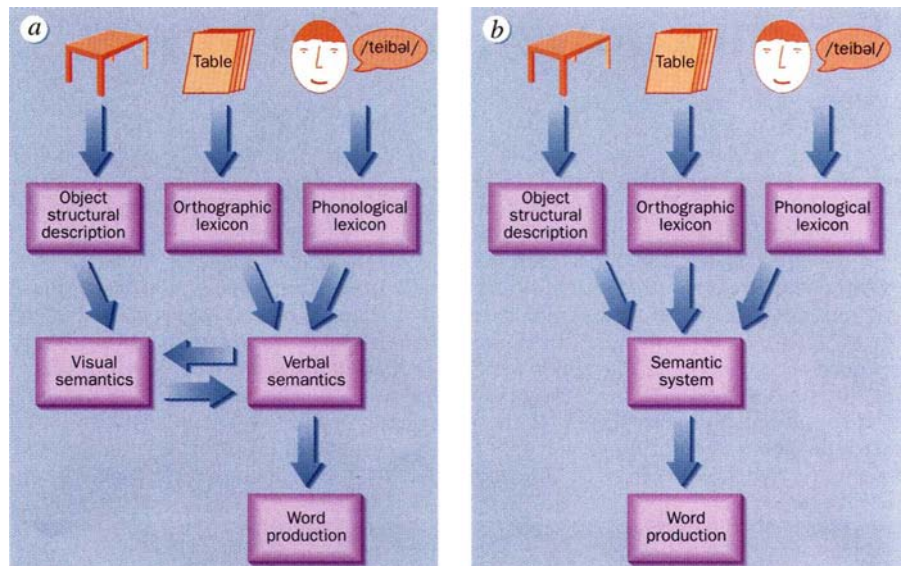
1. Hawking, S. W. *Commun. Math. Phys.* **43**, 199–220 (1975).
2. Bekenstein, J. D. *Phys. Rev. D* **12**, 3077–3085 (1975).
3. Strominger, A. & Vafa, C. *Phys. Lett. B* **379**, 99–104 (1996).
4. Polchinski, J. *Phys. Rev. Lett.* **75**, 4724–4727 (1995).
5. Callan, C. G. Jr & Maldacena, J. M. *Nucl. Phys. B* **472**, 591–610 (1996).
6. Breckenridge, J. C. et al. *Phys. Lett. B* **381**, 423–426 (1996).
7. Maldacena, J. & Strominger, A. *Phys. Rev. Lett.* **77**, 428–429 (1996).
8. Larsen, F. & Wilczek, F. Preprint hep-th/9609084 at xxx.lanl.gov
9. Susskind, L., Thorlacius, L. & Uglum, J. *Phys. Rev. D* **48**, 3743–3761 (1993).
10. 't Hooft, G. in *Salamfest* (eds Ali, A., Ellis, J. & Randjbar-Daemi, S.) 284–296 (World Scientific, Singapore, 1993).
11. Susskind, L. *J. Math. Phys.* **36**, 6377–6396 (1995).

century, has been especially prominent in this context. This disorder involves a selective difficulty in naming visually presented objects, with no corresponding difficulty in naming by touch or in response to verbal definitions. The modality-specific nature of the difficulty cannot be attributed to a deficit in low-level visual processing because the patients can copy the objects they are unable to name, and can even provide accurate mimes for their use.

One interpretation of optic aphasia is that it reflects a disconnection between two semantic systems: one representing visual information and accessed directly only by objects and pictures, the other representing verbal information and accessed only by words^{4,5}. In this framework, the visual semantic system represents the visual properties of objects, for example the fact that tables have legs and a flat surface, and the verbal semantic system represents functional/associative properties of objects, for example the fact that tables are furniture and are used for eating and writing (*a* in the figure). A related claim, also based on the performance of optic aphasics, is that verbal semantics is represented only in the left hemisphere, but visual semantics is represented in both the right and the left hemisphere⁶. However, the neuropsychological evidence is also consistent with the view that semantic knowledge is organized as a functionally (though not necessarily neuroanatomically) unitary system in the left hemisphere, and that it can be accessed directly and independently by both words and visual objects (*b* in the figure)^{7,8}. The significance of the new study reported by Vandenberghe *et al.*¹ is that it provides the evidence needed to choose between the two theories.

The authors measured neural activation, by means of positron-emission tomography, in subjects engaged in semantic processing of pictures and words. One task involved making semantic judgements about visual and the other about functional/associative attributes of concepts. There were three main results: first, there is considerable overlap in the areas activated during semantic processing of pictures and words; second, these areas are distributed widely in the left hemisphere, including regions of the frontal, temporal, parietal and occipital cortices; and third, there are some areas that show activation only in response to pictures or words, but these are not likely to be modality-specific semantic systems because the activation is not specific to a type of semantic content.

This last point is important. The multiple semantics theory predicts that the activation of modality-specific semantic systems depends on the type of information being queried: perceptual for visual semantics, and functional/associative for verbal semantics. But the areas that



Two theories of the relationship between components of the lexical processing system. *a*, The multiple semantics theory, which assumes that the representation of meaning is organized into two modality-specific systems — a visual and a verbal semantic system, each of which is accessed directly only by visual objects and words, respectively. Similarly, language production necessarily goes through the verbal semantic system. The visual semantic system represents only those properties of objects that are acquired through the visual modality; the verbal semantic system represents those properties of objects that are acquired through spoken and written language. In this framework, the neurological disorder optic aphasia is thought to reflect a disconnection between intact visual and verbal semantic systems. *b*, The unitary semantics theory, which assumes that meaning is represented in a functionally unitary system, accessed directly by both visual objects and words. In this framework, optic aphasia is thought to reflect a disconnection between intact neural mechanisms for the recognition of objects and the semantic system.

showed selective activation for type of stimulus (picture or word) did so for both visual and functional/associative judgements. This result shows that the areas selectively activated for pictures and for words are not modality-specific semantic systems but neural mechanisms involved in the recognition of pictures and words, respectively.

Vandenberghe *et al.* show that a functionally unitary semantic system — that is, one that is accessed directly by both words and pictures — is represented in a distributed form in the left hemisphere. But an important question remains unanswered: are the different regions functionally homogeneous, or do they represent different aspects of meaning? This question is especially relevant in the light of neuropsychological evidence showing that brain damage can selectively affect different semantic categories⁹ (for example, difficulty in naming and understanding animal words but not tools or vehicles), as well as functional neuroimaging results showing that different regions of the left hemisphere are activated by different semantic categories^{3,10}. These results indicate that the semantic system is not functionally homogeneous, but its organizing principle remains unclear.

Another issue concerns the relation between semantic knowledge and other aspects of lexical processing. How is the spatially distributed semantic system related to information about word forms

(that is, the sound and spelling of words) and their grammatical properties (for example, the fact that a word is a noun)? Are different parts of the semantic system differentially related to different aspects of lexical knowledge? Here, too, the answers have yet to emerge. Nonetheless, it is clear that we are entering an exciting new phase in the study of the human brain: functional neuroimaging studies are already providing answers to century-old questions and promise to answer increasingly finer-grained questions about the organization of language processes in the brain. □

Alfonso Caramazza is in the Cognitive Neuropsychology Laboratory, Department of Psychology, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138, USA.

1. Vandenberghe, R., Price, C., Wise, R., Josephs, O. & Frackowiak, R. S. J. *Nature* **383**, 254–256 (1996).
2. Caramazza, A. & Zurif, E. B. *Brain Lang.* **3**, 572–582 (1976).
3. Damasio, H., Grabowski, T. J., Tranel, D., Hichwa, R. D. & Damasio, A. R. *Nature* **380**, 499–505 (1996).
4. Beauvois, H. B. & Saffran, E. M. *Brain* **105**, 35–47 (1982).
5. Shallice, T. in *The Cognitive Neuropsychology of Language* (eds Coltheart, M., Sartori, G. & Job, R.) 111–127 (Lawrence Erlbaum, Hove & London, 1987).
6. Coslett, H. B. & Saffran, E. M. *Brain* **112**, 1091–1110 (1989).
7. Caramazza, A., Hillis, A. E., Rapp, B. & Romani, C. *Cogn. Neuropsychol.* **7**, 161–189 (1990).
8. Riddoch, M. J., Humphreys, G. W., Coltheart, M. & Funnell, E. *Cogn. Neuropsychol.* **3**, 3–25 (1988).
9. Warrington, E. K. & Shallice, T. *Brain* **107**, 829–853 (1984).
10. Martin, A., Wiggs, C., Ungerleider, L. & Haxby, J. *Nature* **379**, 649–652 (1996).

Pyramid power, people power

Ian Stewart

How many men does it take to build a pyramid? It's a vexed question, made all the more so by our ignorance of just how the pyramids were built. Herodotus reported that it took 100,000 men to construct the Great Pyramid, but he wrote two thousand years after it was erected, and he wasn't terribly reliable. But we may not need to know *how* the pyramids were built to get a good estimate of the manpower required: the mere fact that they *were* built, in an age not noted for heavy machinery, gives a surprisingly effective

would have had about 23 years — 8,400 days, ignoring stoppages for bad weather, strikes or public holidays — to complete their work.

Wier assumes that the pyramid was built from stone blocks that were quarried, shaped, moved and installed using human muscle-power alone — no doubt via the intermediary of levers, pulleys and other mechanical aids. The pyramid was mainly built in layers from bottom to top. Horizontal transport was on wooden sledges — that much seems highly likely

1,250 men is not an outrageous workforce for an Egyptian king.

For a better estimate, historical detail comes into play. We know where the blocks were quarried, so the terrain and the distance over which they must have been transported are also known quantities. We can even estimate the coefficient of friction of Egyptian sledges. A scene from the tomb of nomarch Djehutihotep at Deir el-Bersha shows a 5-m-high stone statue being pulled along by 172 men. The statue would have weighed some 58 tonnes, so each man was pulling about 330 kg — a coefficient of sliding friction of 0.034, using Wier's assumption that the workers exert a force of about 11.5 kg. Other depictions, and modern experiments with water-lubricated sledges, all lead to a figure of around 0.1.

Putting all this information together, Wier obtains lower and upper estimates of around 9,500 (12,800) men at the start, tailing off to about 2,000 (2,600) when the height reached 100 m, and then dropping rapidly until the final few metres were installed by just 35 (41) men. Alternatively, the workforce could have been constant throughout, with the rate of progress depending on the stage of construction. Men could have been traded between tasks, quarrying, shaping, dragging or installing as necessary. The result this time is between 8,380 and 10,600 men. No doubt the reality was more complex, but for ballpark figures, who cares? As Wier says, "One may suppose that the ancient architects determined the size of the pyramid they wished to construct, or the largest workforce they wished to maintain, then gathered the men and set to work". They had plenty of experience by the time they tackled the really big pyramids, so they should have been able to work out the required numbers.

In broad terms, a workforce of about 10,000 men would have been enough. Since the population of Egypt at the time was probably about 1.1 to 1.5 million, this is less than one per cent of the population. We conclude that there is nothing especially remarkable about pyramid building in terms of manpower; the difficulty was to maintain continuity of organization over periods of several decades. And although Wier does not say so, the calculation also casts doubt on theories that the pyramids were used to soak up surplus labour, to keep the populace in work and avoid political and social unrest. The problem is that they would not have soaked up *enough* — the reduction in the unemployment rate would have been only 2.5 per cent of the (male) labour force. □

Ian Stewart is in the Department of Mathematics, University of Warwick, Coventry CV4 7AL, UK.

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Stone cladding in the suburbs: the Great Pyramid of Khufu (or Cheops) at Giza on the outskirts of Cairo. A big job for the builders, but was the workforce really as large as the 100,000 given by Herodotus?

handle on the question. Or so Stuart Kirkland Wier argues, with a fair degree of conviction, in the *Cambridge Archaeological Journal* (6, 150–163; 1996).

His method is, in effect, a back-of-the-envelope time-and-motion study. Such calculations often put matters into a proper perspective — for example, the manpower requirement of the Apollo Moon-landing was about the same as that used to construct York Minster, a large mediaeval cathedral. Here, the calculations remove much of the mystery from the pyramids, revealing them as a substantial, but entirely plausible, civil engineering project for their time.

Wier takes as his example the biggest pyramid, that of Khufu at Giza, which dates from the 26th century BC. It was originally 146.7 m high, with a square base 230.4 m on a side. Its volume was therefore $2.59 \times 10^6 \text{ m}^3$, and its mass about seven million tonnes. Although it is unlikely that the builders would have known in advance how long Khufu's reign would be, it is reasonable to take the entire reign as the building period. So they

from several sources. We have little idea how vertical transport was performed, but for these purposes the method doesn't greatly matter: what matters is the energetics.

Dividing the volume by the total time, we find that on average the builders had to install 309 m^3 of rock per day. A uniform rate of construction is unlikely, because the taller the structure becomes, the harder it is to lift the rock into place, and the less space there is for men to work. So it seems likely that the installation rate was faster early on, and dropped off later. Wier considers several specific schedules, but the results are much the same in all cases.

The average amount of useful work that a man can perform in a day is about 2.4×10^5 joules — maybe rather more for an experienced worker. The potential energy of the completed pyramid, relative to ground level, is 2.5×10^{12} joules. Therefore, lifting alone would require at least 1.04×10^7 man-days, or 1,250 men over the 8,400-day schedule. This ignores quarrying, shaping and inefficiency. Nevertheless, it puts the issue into perspective:

Bacteria on the rampage

Julian Davies

AWARENESS of the problems raised by the emergence of antibiotic-resistant bacteria is not new — they were anticipated by Sir Alexander Fleming in the 1940s when penicillin became generally available. But today the increased incidence of penicillin-resistant pneumococci is putting the treatment of serious childhood diseases in jeopardy, multidrug-resistant tuberculosis is becoming the norm, and serious diarrhoeal diseases caused by Enterobacteriaceae are often refractory to antibiotic treatment (see table).

How has this situation been allowed to develop? Participants at a recent meeting* expressed intense frustration over the failure to mount a sustained campaign against the spread of resistance, which is an ecological disaster at the environmental, microbial and cellular levels. One note of guarded optimism was that with increasing understanding of the nature, causes and consequences of antibiotic resistance and their relationship to antibiotic use of all kinds, the situation may be checked, if not ameliorated.

Specific recommendations were outlined at the meeting (M. L. Cohen, CDC, Atlanta), and have been proposed on many previous occasions. They include:

- Worldwide surveillance of outbreaks of antibiotic-resistant bacteria.
- More prudent use of antibiotics; revision of clinical practice guidelines.
- New antibiotics; alternative choices for treatment.
- Improved infection control procedures in hospitals.
- Enhanced disease control measures to prevent the emergence of infectious disease; more effective diagnostic procedures.
- The development of new and better vaccines against common microbial diseases.

But implementing them is no easy matter, and will be impossible if the seriousness of the situation is not more widely acknowledged.

In many hospitals it has become difficult to treat the resident (nosocomial) enterococcal strains that can cause life-threatening infections in patients who have undergone invasive surgery. Therapy often relies solely on the use of the glycopeptide antibiotic vancomycin, because the bacteria have become resistant to all other available agents. But vancomycin-resistant enterococci are on the increase, and the worry is that this resistance will be transmitted to the more virulent organism, methicillin-resistant *Staphylococcus aureus* (MRSA).

One might expect that the glycopeptide

antibiotics would have been reserved for these crucial human applications. Not so. For example, avoparcin is chemically related to vancomycin (although its name disguises the fact). In Denmark in 1993, 22 kg of vancomycin were employed in human therapy, while animal use consumed 19,000 kg of avoparcin — inadvertently breaking European Community rules, which state that no agents used in humans and none that cause cross-resis-

mechanisms to rid themselves of extra baggage so as to compete more effectively with their leaner colleagues in the environment. The vast majority of antibiotic-resistance determinants found in pathogenic bacteria are encoded by plasmids, permitting their ready dissemination within the bacterial population, but perhaps at a cost to the host. In comparisons of the growth of antibiotic-sensitive and antibiotic-resistant microbes in the absence of antibiotics, it appears that strains without plasmids do have a slight competitive advantage initially, but the genetic background of the host alters to compensate for the effect of plasmid load.

The most serious antibiotic-resistant bacteria		
Organism	Diseases	Resistance to
Enterobacteriaceae	Bacteraemia, pneumonia, urinary tract and surgical wound infections	Aminoglycosides, beta-lactams, trimethoprim, vancomycin, chloramphenicol
<i>Haemophilus influenzae</i>	Pneumonia, sinusitis, epiglottitis, meningitis, ear infections	Beta-lactams, tetracycline, chloramphenicol, ampicillin, trimethoprim + sulphonamide
<i>Mycobacterium</i> spp.	Tuberculosis	Aminoglycosides, isoniazid, ethambutol, pyrazinamide, rifampin
<i>Neisseria gonorrhoeae</i>	Gonorrhoea	Beta-lactams, penicillins, spectinomycin, tetracycline
<i>Shigella dysenteriae</i>	Severe diarrhoea	Ampicillin, chloramphenicol, trimethoprim + sulphonamide, tetracycline
<i>Pseudomonas aeruginosa</i>	Bacteraemia, pneumonia, urinary tract infections	Aminoglycosides, beta-lactams, tetracycline, chloramphenicol, ciprofloxacin, sulphonamides
<i>Staphylococcus aureus</i>	Bacteraemia, pneumonia, surgical wound infections	Chloramphenicol, rifampin, ciprofloxacin, clindamycin, erythromycin, beta-lactams, tetracycline, trimethoprim
<i>Streptococcus pneumoniae</i>	Meningitis, pneumonia	Chloramphenicol, penicillins, erythromycin
<i>Bacteroides</i> spp.	Anaerobic infections, septicaemia	Penicillins, clindamycin
<i>Enterococcus</i> spp.	Catheter infections, blood poisoning	Penicillins, aminoglycosides, vancomycin, erythromycin, tetracycline

Information from World Health Organization and Centers for Disease Control

tance can be used in animal feed additives. Not surprisingly, resistance to vancomycin sharing the same biochemical mechanism as that found in human isolates is now common in farm animals. Avoparcin was also used in Germany, where vancomycin-resistant enterococci are now widespread and can be detected on supermarket meat products (W. Witte, Robert Koch Inst.). Use of avoparcin in Germany and Denmark is now prohibited, but a powerful lobby is trying to dissuade the European Community from taking general preventative action.

Concerns about the agricultural and aquacultural (mis)use of other classes of antimicrobial agents that have human clinical applications were also raised at the meeting. In particular, the fluoroquinolones, an important class of broad-spectrum antimicrobials, are being used extensively for non-human purposes such as prophylaxis against bacterial infection in fish farming and growth promotion in livestock — with corresponding increases in the population of resistant bacteria.

What about the biology of resistance, for instance the cost/benefit relationship of possessing an antibiotic-resistance gene? It is generally assumed that genetic overload in living organisms is detrimental and, in particular, that microbes have

Mutated hosts then show increased fitness in the absence of antibiotic (R. Lenski, Michigan State Univ.). This work emphasizes the need for a better understanding of the host/plasmid relationship under a variety of conditions. Results from mathematical modelling studies of antibacterial chemotherapy describe various ways that resistance to multiple drugs in bacteria may evolve, and provide pointers to the therapeutic parameters (antibiotic dosage and frequency) that might be adjusted to avoid the development of resistance (B. R. Levin, Emory Univ.).

Impressive analyses of the structures of resistance plasmids isolated from MRSA in Australia have underlined the human health consequences of the molecular plasticity of microbes. An intricate interplay of genetic tricks employing plasmids, transposons, insertion sequences, and interspecies and intergeneric gene capture has been shown to lead to the architectural complexity of staphylococcal plasmids (R. Skurray, Univ. Sydney). The result is a large family of interrelated genetic structures which harbour a variety of different antibiotic-resistance genes; it is likely that the same types of natural genetic engineering have occurred in staphylococcal populations worldwide. (Incidentally, the staphylococci are not alone in showing

*Antibiotic Resistance: Origins, Evolution, Selection and Spread, Ciba Foundation Symposium No. 207, London, 16-18 July 1996.

such enterprise — Australian legal firms advertise their services encouraging patients to sue if they pick up an MRSA infection during a hospital stay.)

The genetic mechanisms by which antibiotic-resistance determinants evolve and are incorporated into bacterial hosts are still poorly understood. But analyses of the structures of resistance-gene clusters in the Enterobacteriaceae have provided a good model for acquisition of resistance through the gene-cassette-integron process. There is now strong evidence for the existence of resistance genes as free open reading frames (cassettes) which recombine into specific expression modules permitting the formation of tandem arrays of resistance genes in the plasmids found in bacterial pathogens; more than 40 such cassettes have been identified (R. M. Hall, CSIRO, Sydney).

Widespread dissemination of resistance genes within the bacterial population is best exemplified by the tetracycline (*tet*) group. This class of antibiotic has been widely used for all imaginable purposes, and the ecology of the *tet* determinants is a model case for resistance-gene dissemination. Clinical and laboratory studies have confirmed the promiscuity of bacterial plasmids, which often employ conjugative transposition to cross what were once thought to be barriers to gene exchange among bacteria (M. C. Roberts, Univ. Washington).

Microbes are formidable competitors for ascendancy on this planet. However, their invisible biodiversity is a fundamental component of all living systems. Given the limitless genetic capacities of microbes, how can current approaches to the treatment of infectious disease be maintained? Antibiotics are essential therapeutic agents, but they must be used judiciously. The many factors involved in both hospital and community medicine are difficult to control. Human demographics and behaviour, together with economic development and changes in land use, influence the spread of infection. The breakdown of public health measures due to famine, poverty and war contributes to the dissemination of resistance, which is compounded by the growth of international travel — antibiotic-resistant microbes are one of the less attractive frequent flyer awards.

So what can be done? The newspaper reports that followed a press conference at the end of the meeting were a one-day wonder. Is there no way to generate sufficient general concern so that antibiotic resistance can be attacked on a continuing basis? □

Julian Davies is in the Department of Microbiology and Immunology, University of British Columbia, 300-6174 University Boulevard, Vancouver, British Columbia, Canada V6T 1Z3.

Timing is everything

Meir Aridor and William E. Balch

THE transport of proteins and lipids in eukaryotic cells is subject to highly selective controls. Vesicular carriers dock at specific target membranes and unload their cargo following membrane fusion. Docking and fusion are distinct events, and occur in response to transient protein-protein interactions.

One class of interactions involves the evolutionarily related members of the

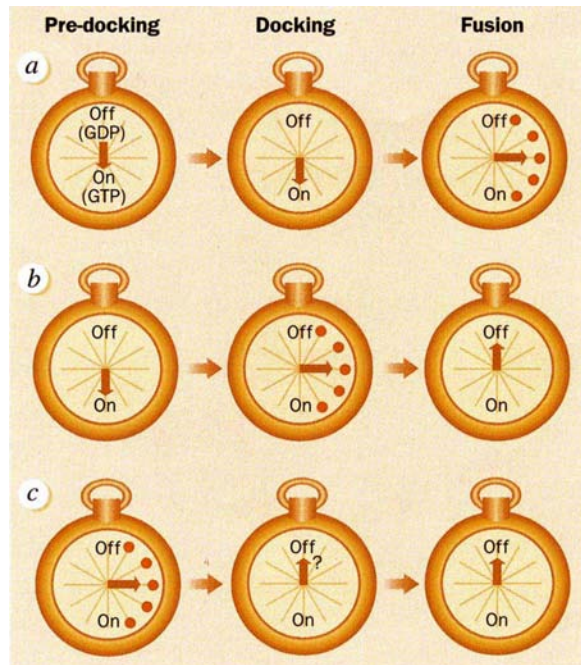
fusion, and the results imply that Rab is a molecular timer that coordinates the protein interactions that occur before docking.

Three models have been proposed to explain the mechanism by which the Rab switch acts in biological membrane fusion. In one of them (*a* in the figure), GTP-bound Rab primes the fusion machinery and controls the fusion event itself. According to this view, activation of GTP hydrolysis is conditional upon fusion.

In the second model (*b* in the figure), the Rab GTPase activity is dependent on a proper docking event, rather than on fusion. Rab detects and coordinates molecular interactions promoting docking. Subsequently, GTP hydrolysis will provide sufficient time for correct protein-complex formation, which will allow fusion to proceed. This model is similar, in principle, to the mode of action of elongation factor EF-Tu in ensuring fidelity of protein translation⁵. One would predict that GTP hydrolysis in this case would depend on productive protein contacts between the vesicle and the target membranes.

In the third model (*c* in the figure), the GTPase activity is independent of both docking and fusion. On activation (GTP binding), Rab-GTP stabilizes the assembly of one or both SNARE complexes. GTP hydrolysis leads to destabilization of the complex, shutting the process off regardless of whether productive docking or fusion has occurred. Regulators of GTP hydrolysis will control events leading to fusion by extending or decreasing the lifetime of such a stabilized state, independently of the initial activating step.

To distinguish between these three possibilities, Rybin *et al.*⁴ used the Rab5 protein, which regulates fusion in the endocytic pathway. Using a Rab5 mutant purified from *Escherichia coli* that binds and hydrolyses xanthosine 5'-triphosphate (XTP) instead of GTP, the XTPase activity of the protein could be followed using an *in vitro* assay that reconstitutes endosome fusion. Using XTP as the source of nucleotide *in vitro*, the authors isolated the Rab5 XTPase switch from endogenous GTPases



Models for the role of the Rab GTPase switch as a molecular timer in controlling biological membrane fusion. *a*, GTP hydrolysis — the timer function converting Rab from the GTP- to GDP-bound form — is conditional upon fusion. *b*, GTP hydrolysis is conditional upon docking. *c*, GTP hydrolysis times the correct assembly of protein complexes on vesicle or target membranes before docking or fusion. The question mark emphasizes the possibility that GTP hydrolysis may be delayed following productive protein-protein interactions culminating in fusion. The work by Rybin *et al.*⁴, discussed here, supports the third model and indicates that Rab acts as a molecular timer that monitors or promotes pre-docking or pre-fusion events.

SNARE machinery, which form separate protein complexes on the vesicle and target membranes^{1,2}. These include the soluble *N*-ethylmaleimide-sensitive factor (NSF), soluble NSF attachment proteins (SNAPs) and their membrane-associated receptors (referred to as SNAREs, for SNAP receptors). The assembly of SNARE complexes is thought to be regulated by a family of small GTPases known as the Rab proteins³, although Rab is not a component of the isolated complexes. Rab acts as a molecular switch that can flip between two conformational states — the active GTP-bound and an inactive GDP-bound form. The paper by Rybin *et al.*⁴ on page 266 of this issue examines the Rab-mediated regulation of events leading to membrane

that are also active in the preparation.

Intriguingly, the results showed that under conditions in which fusion could not occur (in the absence of cytosol), the measured rate of XTP hydrolysis was no different from that measured under fusion-promoting conditions. These results negate the first model, as they show that the onset and timing of the Rab GTPase switch is uncoupled from the fusion process itself. Even more surprisingly, the Rab GTPase activity does not require docking (*b* in the figure). Dilution of the membranes in the assay to reduce their ability to interact with each other — which would be expected to markedly affect docking — also had no effect on the rate of nucleotide hydrolysis. These results therefore support the third model, in which Rab is a timer that monitors or promotes pre-docking/pre-fusion events.

This hypothesis has implications for understanding the general protein interactions leading to membrane fusion. Proteins that bind and stabilize the GTP-bound form of Rab, such as Rabaptin⁶, may prolong the lifetime of the competent state and thereby promote events leading to fusion. Moreover, if the frequency and lifetime of the assembled complex are critical for function, as proposed by Rybin *et al.*⁴, overexpression of SNARE components should promote vesicle fusion independently of Rab. This has indeed been observed in yeast⁷. Ypt1p, the yeast homologue of the Rab1 protein, is essential for the interaction between the yeast SNARE proteins Bos1p and Sec22p, found on vesicles derived from the endoplasmic reticulum. Overexpression of both Bos1p and Sec22p can compensate for the loss of the Ypt1p.

Finally, the recruitment and release of SNAP occurs before completion of Rab function⁸. This is consistent with evidence that NSF is involved in a pre-fusion event⁹. Although we cannot rule out the possibility that downstream events may still be controlled by Rab (the question mark in *c* in the figure), docking and fusion could rely on protein interactions that follow NSF/SNAP and Rab, and

therefore remain to be discovered.

The properties ascribed to Rab shed light on the general properties of the fusion process itself. Fusion is a dynamic event that is reversible and unstable. For example, when the exocytic fusion of secretory granules is measured directly in patch-clamped mast cells, there is an initial opening of a fusion pore¹⁰. This non-selective channel opens and closes in a process described as 'flickering', although fusion does not go to completion. When sufficiently stabilized, the pore dilates completely and fusion occurs. Therefore, non-hydrolysable GTP analogues — which maintain GTPases in the active state — promote fusion-pore formation and extensive exocytosis. The involvement of Rab in these events has been documented^{11–13}.

Although the molecular mechanism and mode of regulation of the final fusion step are unknown, the results of Rybin *et*

*al.*⁴ indicate that the triggering of regulated exocytosis through cell-surface receptors leading to GTPase activation may stabilize the machinery found on each membrane, thereby promoting fusion. In the constitutive exocytic and endocytic pathways, different Rab timers would stabilize the assembly of components found on distinct vesicles and compartments required for docking and/or fusion. It is therefore the frequency (regulated by GTP binding through the activity of guanine nucleotide exchange factor) and the duration (regulated by GTP hydrolysis through the action of the GTPase-activated protein) that underscores events culminating in fusion. □

Meir Aridor and William E. Balch are in the Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

EXO BIOLOGY

Dusting off panspermia

Paul Parsons

MANY people find it fascinating that life has been able to evolve on Earth. More fascinating still is the idea that life on Earth originated elsewhere in the Galaxy, and was transmitted to our planet by space-borne microorganisms — the panspermia hypothesis¹. Eleven years ago, a study showed that it may be possible to transport life over large interstellar distances², but until now none of the various manifestations of the theory has been able to propel life beyond the gravity of an organism's home star system. In the August issue of the *Journal of the Royal Astronomical Society of Canada*³, and in earlier work⁴, Secker, Lepock and Wesson refine an intriguing variant of the panspermia idea. They show that it is possible, and even predict the rate of propagation.

Panspermia was first proposed early this century by Arrhenius⁵. Since then, a number of different incarnations of the theory have appeared in the literature. Lithopanspermia⁶ is the distribution of microbes by meteorites, and has enjoyed recent fame. Directed panspermia⁷, which proposes our deliberate deposition on Earth by some extraterrestrial intelligence, is perhaps more fanciful. Secker *et al.* concentrate on the third possibility: radiopanspermia, or the dissemination of organisms by starlight.

Stellar radiation can exert a substantial pressure on nearby matter. The astrophysicist Arthur Eddington showed that stars of more than about 60 solar masses cannot exist, because they would be torn apart by radiation pressure. Radiopanspermia uses a star's radiation pres-

sure to propel individual microorganisms into interstellar space.

But, chiefly because of the damaging ultraviolet radiation emitted by stars, the life expectancy of bare organisms in a stellar neighbourhood 50 times the size of the Earth's orbit is only a few days. Secker *et al.* overcome this problem by supplementing the radiopanspermia theory with one of the characteristics of lithopanspermia — shielding. They argue that by encapsulating bacteria or viruses in small protective mantles of material, roughly half a micrometre thick, it is possible to screen them from the lethal radiation.

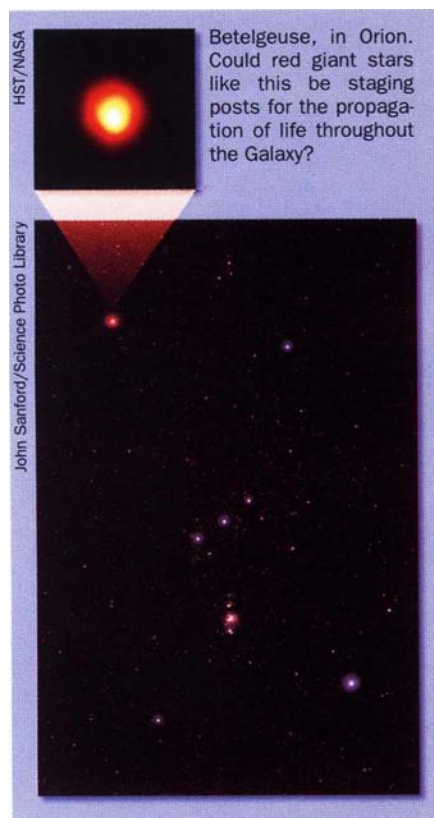
What could such a mantle be composed of? Ice was rejected because of its poor opacity to the most damaging ultraviolet wavelengths. The next choice was carbon. Encasing organisms inside a grain of dust, with a composition matching that found in type-I carbonaceous chondrite meteorites, renders them resilient enough to withstand stellar ultraviolet radiation. The extra mass of such a dust layer, however, means that the radiation pressure must be intense in order to accelerate each organism to escape velocity from a starting radius comparable to the Earth–Sun distance.

The most promising sources of such intense radiation are red giants. These are highly evolved stars, hundreds of times the size of the Sun, and hundreds of thousands of times more luminous. They are bright enough to carry organisms into space, and, being cool, the light they emit is relatively red, with little ultraviolet content. They are also by definition old, and are therefore more likely to harbour life

1. Söllner, T. *et al.* *Nature* **362**, 318–324 (1993).
2. Bennett, M. K. & Scheller, R. H. *Proc. Natl Acad. Sci. USA* **90**, 2559–2563 (1993).
3. Nuoffer, C. & Balch, W. E. *Annu. Rev. Biochem.* **63**, 949–990 (1994).
4. Rybin, V. *et al.* *Nature* **383**, 266–269 (1996).
5. Burgess, S. M. & Guthrie, C. *Trends Biochem. Sci.* **18**, 381–384 (1993).
6. Stenmark, H., Vitale, G., Ullrich, O. & Zerial, M. *Cell* **83**, 423–432 (1995).
7. Lian, J. P., Stone, S., Jiang, Y., Lyons, P. & Ferro-Novick, S. *Nature* **372**, 696–701 (1994).
8. Mayer, A., Wickner, W. & Haas, A. *Cell* **85**, 83–94 (1996).
9. Banerjee, A., Barry, V. A., DasGupta, B. R. & Martin, T. F. J. *J. Biol. Chem.* **271**, 20223–20226 (1996).
10. Lindau, M. & Almers, W. *Curr. Opin. Cell Biol.* **7**, 509–517 (1995).
11. Fischer von Mollard, G., Stahl, B., Li, C., Sudhof, T. C. & Jahn, R. *Trends Biochem. Sci.* **19**, 164–168 (1994).
12. Oberhauser, A. F., Monck, J. R., Balch, W. E. & Fernandez, J. M. *Nature* **360**, 270–273 (1992).
13. Richmond, J. & Haydon, P. G. *FEBS Lett.* **326**, 124–130 (1993).

on their surrounding planets than are young, massive stars (which form the other class of stars that would be luminous enough).

Secker and collaborators estimate the survival chances of radiation-driven organisms during interstellar transit. They find that the physical size of viruses and bacteria is of primary importance: smaller organisms can support a thicker dust shield while remaining light enough to be driven by radiation pressure, and so they



are more likely to survive. Without shielding, all the organisms they consider are killed before reaching the relative safety of interstellar space.

But even that is no haven. Although it lacks the intense ultraviolet radiation found in stellar neighbourhoods, interstellar space is bathed in diffuse radiation that can eventually destroy microorganisms. Because of the coldness of interstellar space, irradiated strands of DNA or RNA cannot repair themselves, and the damage they sustain is cumulative. But screening by a carbonaceous mantle allows all the considered strains to survive their perilous interstellar journey.

Radiation damage is not the only danger that organisms face. Violent phenomena such as volcanic activity and meteorite impacts are thought to be the most likely ways for microorganisms to leave their home planet⁸, and these events inevitably generate shock waves in the planet's atmosphere, destroying some of the organic material as it tries to leave. Entry into the atmosphere of the new host

planet poses another threat to travelling viruses and bacteria⁹. The Sun's radiation can only partially decelerate particles cast out by large red giant luminosities, so most of them will encounter the Earth's atmosphere at high speed.

In order to expose the organisms to their new environment, their dust jackets must be fractured by atmospheric drag during entry, but too much heat will destroy them. Fortunately, there is likely to be a wide range of speeds — Secker and colleagues estimate that the relative velocity of incoming radiation-driven particles with respect to the Earth will vary between 0 and 50 km s⁻¹. Some fraction of Earth-bound organisms will then be slow enough to be allowed safe passage into the planet's atmosphere.

But is radiopanspermia really plausible? A high proportion of radiation-driven, dust-clad organisms, could travel 20 light years in about a million years. Beyond that, interstellar radiation and cosmic-ray particles become a serious threat. To travel farther, microorganisms would need to establish life around a new star, which in time could cast out new organisms, and so on. Life propagating in this way would have had time to populate a large fraction of the Galaxy.

The group point out that even dead bacteria could accelerate the emergence of life on a planet. Provided the environment is conducive, biological material such as DNA fragments would be available for replication, and could greatly improve the chances of life evolving.

To verify the radiopanspermia concept experimentally, we would have to detect live organisms in interstellar space. That seems unlikely by direct means¹, although the detection of apparently pre-solar extraterrestrial helium inside fullerenes¹⁰ at the Sudbury meteorite impact site is encouraging. If the theory is correct, the notion of life existing in isolated pockets in the Galaxy is a misleading one. It becomes more appropriate to think of a ubiquitous 'galactic life', of which life on Earth is just a small part. □

Paul Parsons is in the Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK.

1. Hornebeck, G. *Planet. Space Sci.* **43**, 189–217 (1995).
2. Weber, P. & Greenberg, J. M. *Nature* **316**, 403–407 (1985).
3. Secker, J., Wesson, P. S. & Lepock, J. R. *J. R. Astron. Soc. Can.* **90**, No. 4, 184–192 (1996).
4. Secker, J., Lepock, J. & Wesson, P. *Astrophys. Space Sci.* **219**, 1–28 (1994).
5. Arrhenius, S. *Worlds in the Making* (Harper & Row, New York, 1908).
6. Tobias, C. A. & Todd, P. *Space Radiation Biology and Related Topics* (Academic, New York, 1974).
7. Crick, F. H. C. & Orgel, L. E. *Icarus* **19**, 341–346 (1973).
8. Melosh, H. J. *Nature* **332**, 687–688 (1988).
9. Hornebeck, G. & Bucker, H. *Origins of Life* **16**, 414–415 (1985).
10. Becker, L., Poseda, R. J. & Bada, J. L. *Science* **272**, 249–252 (1996).

DAEDALUS

Silence overhead

THE big trouble with the jet engine is its noise. This is largely generated by turbulent shear when the high-speed turbine exhaust erupts into the still air around it. The exhaust from a turbofan engine emerges into moving fan-blown air, reducing this shear and its noise. A bolder plan would surround the engine with microphones. They would pick up its noise, and feed it to amplifiers and speakers on the aircraft. These would re-radiate it, but in anti-phase. Noise and anti-noise would cancel each other out, giving blessed silence.

A snag with this idea is that no conventional loudspeaker could possibly make as much noise as a jet engine. To put out enough sonic power, the whole fuselage itself would have to be vibrated. Daedalus has a neater notion.

He recalls the ionophone, a loudspeaker with no moving parts. It passes a current through ionized air in a magnetic field, inducing loudspeaker action in the air itself. A jet exhaust is already slightly ionized from the combustion. Load the fuel with a caesium compound, and the exhaust will conduct more strongly; beam in ionizing radiation from an X-ray tube or radioactive source, and it will conduct more strongly still. It should then be possible to pass an audio current through the exhaust. Put a set of magnets around the tail-pipe, and the jet would 'speak' that sound.

So DREADCO engineers are fitting the tail-pipes of a small executive jet with microphones and electrodes, and are devising electronics to accept the input noise from the former, and cancel it by anti-noise created by feedback currents delivered to the latter. With no speaker cones to be vibrated, kilowatts of audio power could be deployed. Since the original noise and the cancelling anti-noise will be generated at exactly the same place, perfect cancellation seems possible; but Daedalus will be happy if the first trials reduce the racket by even a few decibels. The system will then be optimized towards the unattainable ideal of utter monastic silence. The engine will no longer be wasting energy in useless noise, so it should even deliver more thrust.

DREADCO's 'Mute-a-jet' system should transform our skies and airports. But these days, while noise is annoying, silence is unendurable. Daedalus is already modifying the electronics to superimpose a chosen additional sound on a jet quietened by his system. A plane could then radiate cheerful pop music as it took off, or dispense from the skies commercials for consumer products, or snatches of vacuous chatter from some popular talk show.

David Jones

Origin of ant soldiers

SIR — For more than 100 years, biologists have believed that ant social organization required an infertile worker caste from which, in some species, a special subcaste of soldiers had been selected. Here we suggest that ant soldiers form a distinct caste which originated directly from reproductive females (gynes) by modifying their development in a way parallel to, but independent from, that of workers.

The only hypothesis for the origin of ant soldiers was formulated in 1894 by Emery¹. It assumes that ant soldiers originate by means of the following stepwise process: (1) presence of polymorphic workers; (2) worker polymorphism becomes bimodal; and (3) intermediates between the two modes disappear and morphologically stable workers and soldiers remain. Later, the same hypothesis was re-proposed by Wilson^{2,3}.

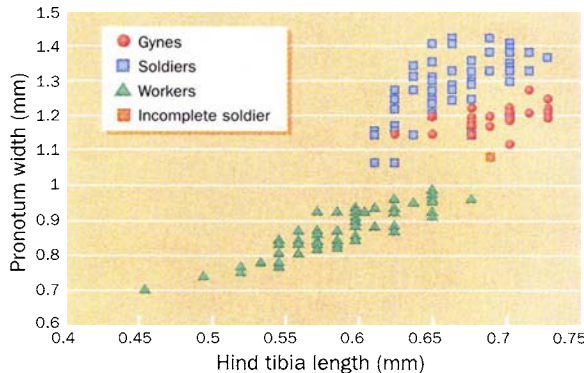
The Emery–Wilson conjecture assumes that workers and soldiers are not separate castes with different origins but simple subsets (subcastes) of a biologically homogeneous worker caste. This concept is made clear by Wilson² who states that “subcastes in a polymorphic worker series are designated approximately according to the relative size of the allometric organ as the *major* (soldier), *media*, and *minor*”. This view has influenced the field to the point that the term “soldier”, established since 1852 (ref. 4) is sometimes entirely banished in favour of the phrase “major worker”.

We propose here that ant soldiers are derived from gynes, not from workers. Our hypothesis is supported by evidence from morphology, behaviour, physiology and pathology. Visual examination of workers, soldiers and gynes of the same species of ants suggests that soldiers are much more similar to gynes than are workers. A quantitative appreciation of this fact results from the conventional, biometric approach to caste description^{2,5} when gynes are also considered (see figure). The soldier–gyne similarity is at its greatest in some species of the genera *Zacryptocerus* or *Camponotus* (*Colobopsis*) where the soldier caste is specialized for a particular task: blocking the nest entrance⁶. These soldiers reveal exorbitant morphological adaptations appropriate to their function, like the head in the form of a shield or a plug. Gynes of these species bear the same, probably non-adaptive, shield-shaped or plug-shaped head⁶.

Ant workers exhibit a broad behavioural repertoire, as opposed to a very narrow one for soldiers and gynes. In *Camponotus* (*Colobopsis*) spp. workers, soldiers and

gynes perform respectively 28, 14 and 5 different behavioural acts; in *Z. varians* the numbers are 31, 2 and 2 (ref. 7).

In two species of *Pheidole*, soldiers bear wing disks (as do gynes) until the pupal stage, whereas no trace of such disks can be detected among workers^{8,9}. Both soldiers¹⁰ and gynes¹¹ of *Pheidole* are determined during a methoprene-sensitive phase; in *C. aethiops*, while gynes and soldiers develop exclusively from hibernating



Bivariate scattergram of two body measures (maximum pronotal width and hind tibia length) of a random sample of 62 workers, 61 soldiers, a unique incomplete soldier and all available gynes (38) from a single nest of *Z. angustus* (Mayr) from Botucatu (São Paulo, Brazil). The incomplete soldier lies within the soldier–gyne biometric range, whereas normally developed soldiers reside much closer to gynes than workers.

larvae, workers also develop from the summer brood¹².

Queenless soldiers of three *Camponotus* species lay haploid eggs while workers remain sterile^{13–15}. The difference in fertility between *Camponotus* (*Colobopsis*) soldiers and workers is further confirmed by differences both in the size of the ovaries and in the number of ovarioles^{16,17}. We dissected a random sample of 10 soldiers and 10 workers from a queenless colony of *Z. clypeatus*: 9 soldiers showed large ovaries containing one or more large oocytes or an egg while all workers had smaller ovaries, though 4 worker ovaries contained a small, probably trophic, egg in their distal portion. Soldiers are, therefore, more fertile than workers, like gynes.

In *P. pallidula*, nematodes may prevent individual ants from developing completely

into soldiers: these “failed soldiers” show gyne traits such as ocelli and mesosomal morphology not present in workers¹⁸.

A consequence of considering soldiers as a subcaste of workers is the need to identify the “primitive caste” for species with more than one subcaste. The literature suggests criteria with which to identify this hypothetical “primitive caste”¹⁹. As a result of this erroneous premise, some researchers reached the logically correct but biologically improbable conclusion that the primitive condition for certain ant species should be a workerless society of soldiers²⁰.

It seems unlikely that ants produced gyne-diverging workers to re-select among them for gyne-converging soldiers. Soldiers would have been produced in a much more parsimonious way directly from gynes. Parsimony represents an excellent criterion for decision making as long as there are no external arguments contradicting it. Our examples are drawn from the sole three ant genera for which detailed information is available out of about 15 with dimorphic neuters. We do not believe that the largest individuals of continuously polymorphic species are true soldiers. Soldiers originated several times in the course of ant evolution (as demonstrated by their presence in phylogenetically distant genera) but the morphological similarity between soldiers and gynes appears to be the rule when two discrete neuter castes are present.

We are unable to identify any biological or other counterarguments supporting the Emery–Wilson hypothesis. Rather, direct production of soldiers from gynes, without the unnecessary intermediate worker step, seems highly likely in all cases in which true soldiers exist as a discrete caste.

Cesare Baroni Urbani

Zoologisches Institut der Universität,
Rheinsprung 9,
CH-4051 Basel, Switzerland

Luc Passera

Laboratoire d’Ethologie & Psychologie
Animale,
UMR-CNRS 5550, Université Paul Sabatier,
118 Route de Narbonne,
31062 Toulouse Cedex, France

- Emery, C. *Biol. Zentralbl.* **14**, 53–59 (1894).
- Wilson, E. O. *Q. Rev. Biol.* **28**, 136–156 (1953).
- Wilson, E. O. *Insectes Soc.* **1**, 75–80 (1954).
- Heer, O. *Zürcher. Jugend* **54**, 1–24 (1852).
- Huxley, J. S. *Biol. Zentralbl.* **47**, 151–163 (1927).
- Wheeler, W. M. *Ants* (Columbia Univ. Press, New York, 1910).
- Cole, B. J. *Insectes Soc.* **27**, 265–275 (1980).
- Wheeler, D. E. & Nijhout, H. F. *Int. J. Insect Morphol. Embryol.* **10**, 131–139 (1981).
- Suzzoni, J.-P., Passera, L. & Strambi, A. in *La Communication Chez les Sociétés d’Insectes* (eds de Haro, A. & Espadaler, X.) 147–156 (1982).
- Wheeler, D. E. & Nijhout, H. F. *J. Insect Physiol.* **29**, 847–854 (1983).

- Passera, L. & Suzzoni, J.-P. *C. R. Acad. Sci. D* **287**, 1231–1233 (1978).
- Dartigues, D. & Passera, L. *Bull. Soc. Zool. Fr.* **104**, 197–207 (1979).
- Bier, K. *Zool. Anz.* **150**, 283–288 (1953).
- Dartigues, D. & Passera, L. *Ann. Soc. Entomol. Fr.* **15**, 109–116 (1979).
- Benois, A. thesis, Univ. Toulouse (1969).
- Wilson, E. O. *Psyche* **81**, 182–188 (1974).
- Hasegawa, E. *Evolution* **48**, 1121–1129 (1994).
- Passera, L. *Insectes Soc.* **23**, 559–575 (1976).
- Wilson, E. O. *Behav. Ecol. Sociobiol.* **7**, 143–146 (1980).
- Wheeler, D. E. *Psyche* **91**, 171–192 (1984).

Mitochondria and male disease

SIR — Mitochondrial mutations are associated with a wide range of degenerative conditions, including blindness caused by Leber's hereditary optic neuropathy, and various muscle and heart conditions in maternally inherited myopathy and cardiomyopathy. Recent data suggest an even broader array of maladies associated with mitochondrial mutations and poor aerobic performance, from Parkinson's and Alzheimer's diseases to reduced sperm motility and male infertility (reviewed in ref. 1).

Mitochondria are transmitted via the mother in most organisms. This matrilineal inheritance creates an important and previously unnoted male–female asymmetry in the expected severity of mitochondrial disease. Consider a germline mitochondrial mutation with severe effects on males but only mild effects on females. This mutation could increase to relatively high frequency because natural selection of mitochondria occurs only in females. Males do not transmit mitochondria, and therefore male-specific phenotypes have no fitness consequences for mitochondria.

Simple models of population genetics describe the quantitative effects of this male–female asymmetry. The equilibrium frequency of germline mitochondrial mutations is approximately $q = \mu/s_f$, where μ is the mitochondrial mutation rate in the female germ line, and $1-s_f$ is the fitness of a female with the mutation relative to a normal female with a fitness of one. A male carrying the mitochondrial mutation has a relative fitness of $1-s_m$ when compared with a normal male with a fitness of one. For example, if a mutation arose with frequency $\mu = 10^{-4}$, with severe effects on sperm motility and a reduction in male fertility by one-half ($s_m = 0.5$) but only mild effects on females ($s_f = 0.01$), then $q = \mu/s_f = 0.01$. This mutation, occurring at an equilibrium frequency of one per cent of the population, would have strong fitness consequences for males but negligible effects on females.

At least one maternally inherited mitochondrial disease affects males more severely than females. Approximately 85 per cent of individuals affected by Leber's hereditary optic neuropathy are male¹. Several distinct base-pair substitutions in germline mitochondrial DNA have been directly associated with this disease. Linkage analysis suggested that a deleterious

X-linked mutation in combination with defective mitochondria may be the cause of the sex bias². However, several further studies failed to support a model of X-linkage (see, for example, ref. 3). An analysis of X-inactivation also argues against X-linkage as the cause of sex-bias⁴. Furthermore, this particular neuropathy typically has an earlier age of onset in males than females. Present evidence therefore suggests that males are more susceptible to the mitochondrial defects.

In another example, an inherited mitochondrial mutation was associated with early-onset Pearson marrow–pancreas syndrome in a male (high s_m) but only late-onset, progressive eye disease in the mother (low s_f)⁵. Finally, a mother and her two sons shared a tendency for spontaneous, somatic deletions in mitochondrial DNA⁶. The mother was asymptomatic but the two sons were severely affected by anaemia and mitochondrial myopathy. This suggests that males may be more susceptible than females to certain mitochondrial defects.

Male–female dimorphism of mitochondrial effects is possible in various organs, such as the heart. The difference between the sexes is likely to be small, but perhaps significant. Ascertainment will be difficult because there is undoubtedly a wide spectrum of mutations, many with highly correlated effects on the sexes (high correlation between s_m and s_f). However, natural selection creates an asymmetric sieve, pushing down in frequency any mutations with high s_f values but allowing through mutations with high s_m but low s_f . This selective sieve could explain some aspects of male–female dimorphism in heart disease and other degenerative syndromes.

Sperm vigour and male fertility are traits that deserve close study. Sperm obtain energy for motility from a densely packed group of mitochondria at the base of the flagellum⁷. Almost all energy produced by this pack of mitochondria is used to drive the flagellum and propel the sperm. Reduced power output translates directly into reduced motility and probably into reduced fertility. This tight coupling between capacity and performance differs from most tissues, which function reasonably well when mitochondrial capacity is reduced by as much as 80 or 90 per cent. Disease is observed only when mitochondrial function drops below a threshold of 10 or 20 per cent of full capacity¹.

Aberrations in sperm motility and male fertility are widely observed in humans and other species⁸. Such traits are puzzling, given the powerful selection against alleles with deleterious effects on fertility. Sperm dysfunction may be explicable, however, if mitochondrial mutations arise

that strongly affect sperm vigour but have only weak effects on female fitness. Both somatic mitochondrial DNA deletions^{8,9} and inherited mitochondrial disease¹⁰ have been associated with reduced sperm motility and poor fertility. Whether inherited mutations are sufficiently frequent to be an important cause of infertility will depend on the correlated fitness effect of these mutations when in females (the correlation between s_f and s_m).

The logic of the population genetics argument is indisputable. The important questions are how frequently mutations with $s_m > s_f$ arise, and to what extent male infertility or excess mortality can be explained by such mutations. Rapid progress on the mitochondrial genetics of humans and mammalian models will provide some clues in the near future. Experiments and screens of natural populations of *Drosophila* can supplement the mammalian data. Mutagenesis experiments could provide data on the correlation between s_m and s_f for new mutations, with particular attention to sperm dysfunction. The idea of a selective sieve could be examined by comparing the distribution of s_m and s_f in new mutations relative to the distribution after selection. These model systems can also be used to study how frequently compensatory nuclear changes mask male-specific mitochondrial defects.

Male–female dimorphism of mitochondrial disease is a subset of a larger problem. Many symbiotic microorganisms are transmitted matrilineally¹¹. The function of these symbionts in males varies widely. The broad problem can be described as the competence of uniparentally inherited symbionts in the nontransmitting sex. Once viewed in this way, many interesting puzzles emerge.

S. A. Frank

Department of Ecology and Evolutionary Biology,

University of California,
Irvine, California 92697-2525, USA

L. D. Hurst

Department of Genetics,
University of Cambridge,
Cambridge CB2 3EH, UK

- Wallace, D. C. *Annu. Rev. Biochem.* **61**, 1175–1212 (1992).
- Vilkkki, J. et al. *Am. J. Hum. Genet.* **48**, 486–491 (1991).
- Juvonen, V., Vilkkki, J., Aula, P., Nikoskelainen, E. & Savontaus, M. L. *Am. J. Hum. Genet.* **53**, 289–292 (1993).
- Oostra, R. J., Kemp, S., Bolhuis, P. A. & Bleekerwagemakers, E. M. *Hum. Genet.* **97**, 500–505 (1996).
- Bernes, S. M. et al. *J. Pediatr.* **123**, 598–602 (1993).
- Casademont, J. et al. *Hum. Mol. Genet.* **3**, 1945–1949 (1994).
- Cardullo, R. A. & Baltz, J. M. *Cell Motil.* **19**, 180–188 (1991).
- Kao, S.-H., Chao, H.-T. & Wei, Y.-H. *Biol. Reprod.* **52**, 729–736 (1995).
- Cummin, J. M., Meloni, B. P. & Jaquier, A. M. in *25th Annual Conference of Australian Society for Reproductive Biology*, Dunedin, NZ, abstr. 1 (1993).
- Folgerø, T. et al. *Hum. Reprod.* **8**, 1863–1868 (1993).
- Buchner, P. *Endosymbiosis of Animals with Plant Microorganisms* (Wiley, New York, 1965).

Scientific Correspondence

Scientific Correspondence is intended to provide a forum in which readers may raise points of a scientific character. Priority will be given to letters of fewer than 500 words.

Evolutionary fixation of RNA editing

SIR — Gene sequences encoded in the organellar genomes of many organisms are incompatible with function of the gene product. The 'errors' in the sequence are corrected after transcription of the gene by processes described by the term 'RNA editing'¹. The persistence throughout evolution of organellar RNA editing is perplexing, because it might have been expected that back-mutations would have quickly removed the need for RNA editing systems.

In the mitochondrial genome of marsupials, the transfer (t)RNA^{Asp} gene carries the anticodon GCC, which potentially recognizes two glycine codons². The second position of the anticodon (position 35) is changed post-transcriptionally so that the canonical aspartate anticodon GUC is created in approximately 50% of the tRNA pool^{2,3}. Considering that this seems to be the only edited position in this genome⁴, and in view of the high rate of evolution of the mammalian mitochondrial genome⁵ that should allow for frequent back-mutations, it is particularly puzzling why this nucleotide has come to be corrected by an RNA editing event rather than by a simple substitution in the gene.

It is known that whereas the edited version of the tRNA^{Asp} (anticodon GUC) gets aminoacylated with aspartate, its

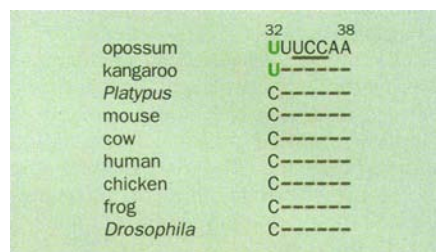


FIG. 1 Alignment of anticodon loop sequences (nucleotides 32–38) of the tRNA^{Gly} from two marsupials (ref. 4 and M. Dörner, personal communication) and some representative animals^{8,12}. The U at position 32 is green; identical bases are indicated by a dashed line.

unedited version (anticodon GCC) is aminoacylated with glycine⁶, suggesting that the unedited tRNA^{Asp}(GCC) functions as a glycyl-tRNA. At the same time, a functional tRNA^{Gly} with the expected anticodon UCC exists in marsupial mitochondria^{4,6}. Whereas the tRNA^{Gly}(UCC) will recognize all four glycine codons (GGN)⁷, the unedited version of tRNA^{Asp}(GCC) will recognize only two of these (GGY). Considering that the mitochondrial translation machinery of animals contains a limited set of only 22 tRNAs, it is surprising that two tRNAs recognizing an overlapping set of glycine codons should exist in marsupials.

A comparison of the marsupial

tRNA^{Gly}(UCC) gene sequences with those of other animals⁸ (Fig. 1) shows that at position 32, two nucleotides 5' of the anticodon, marsupials carry a U residue whereas most other metazoans have a C residue. In an *Escherichia coli* translation system, it has been shown that a C-to-U substitution at that position restricts the decoding capacity of tRNA^{Gly}(UCC) to the codons GGA and GGG⁹. Thus, it is likely that the marsupial tRNA^{Gly}(UCC) decodes only the codons GGA and GGG, and that the remaining glycine codons GGC and GGU are read by the unedited tRNA^{Asp}(GCC). Consequently, the decoding capacities of the two glycyl-tRNAs would not overlap (Fig. 2).

These findings suggest the following evolutionary pathway for the fixation of RNA editing in marsupial mitochondria: the first mutation was a T-to-C transition in the anticodon of tRNA^{Asp}. Because this mutation would have been lethal for the organism, a 'precursor' editing activity probably existed. Such an activity could have been related to, for example, the de-amination of other biomolecules¹⁰. The adaptation of this activity to the editing of tRNA^{Asp} was presumably made possible by the multi-copy state of the mitochondrial genome, which can allow the selection of adaptive changes to take place. At first, the function of the unedited form of tRNA^{Asp} as a glycine isoacceptor would have been redundant. But it would have then created a situation where the substitution at position 32 of the tRNA^{Gly}(UCC) could occur, limiting its decoding capacity to the two glycine codons not decoded by the unedited form of tRNA^{Asp}(GCC).

Once fixed in the germ line, the secondary substitution in the tRNA^{Gly}(UCC) gene would have made an elimination of the tRNA^{Asp} substitution by a back-mutation lethal, as such a mutation would leave two out of four glycine codons without a decoding tRNA (Fig. 2). Thus, the elimination of the editing would require two back-mutations in the same mitochondrial lineage. In effect, the secondary mutation in the tRNA^{Gly} has therefore caused the editing of tRNA^{Asp} to become fixed in marsupial mitochondria.

We suggest that such combinations of mutations represent 'genetic gridlocks' that could be responsible for the occurrence of RNA editing in organellar

genomes. Because in other systems many positions have become dependent on RNA editing subsequent to the initial events, the initial mutations that have caused the evolutionary fixation of editing are probably indiscernible today.

An additional fascinating aspect of this system is that the primary transcript of the marsupial tRNA^{Asp} gene represents a naturally occurring tRNA that, depending on its state of editing, is assigned to one or the other codon

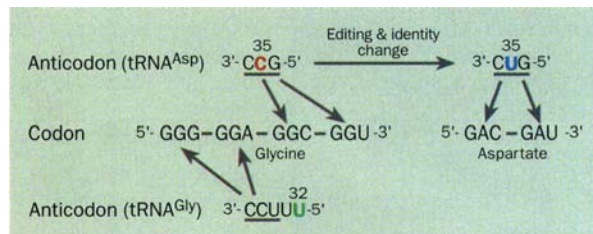


FIG. 2 Schematic illustration of the putative functions of tRNA^{Gly} and tRNA^{Asp} in marsupials. The tRNA^{Gly}(UCC) is restricted to two glycine codons owing to the substitution at position 32. The remaining two glycine codons can be recognized by the unedited version of tRNA^{Asp}(GCC). Subsequent to anticodon editing, this tRNA recognizes the two aspartate codons.

family. When such cases of RNA editing of anticodons are not accompanied by a change in tRNA identity, they may have been instrumental in the reassignments of codons to new amino acids in mitochondria¹¹.

G. Valentin Börner

Svante Pääbo

*Institute of Zoology,
University of Munich,
PO Box 202136,
D-80021 Munich,
Germany*

- Benne, R. *Curr. Opin. Genet. Dev.* **6**, 221–231 (1996).
- Janke, A. & Pääbo, S. *Nucleic Acids Res.* **21**, 1523–1525 (1993).
- Mörl, M., Dörner, M. & Pääbo, S. *Nucleic Acids Res.* **23**, 3380–3384 (1995).
- Janke, A., Feldmaier-Fuchs, G., Thomas, W. K., von Haeseler, A. & Pääbo, S. *Genetics* **137**, 243–256 (1994).
- Brown, W. M., George, M. & Wilson, A. C. *Proc. Natl Acad. Sci. USA* **76**, 1971–1976 (1979).
- Börner, G. V., Mörl, M., Janke, A. & Pääbo, S. *EMBO J.* (in the press).
- Barrell, B. G. et al. *Proc. Natl Acad. Sci. USA* **77**, 3164–3166 (1980).
- Steinberg, S., Misch, A. & Sprinzl, M. *Nucleic Acids Res.* **21**, 3011–3015 (1993).
- Lustig, F. et al. *Proc. Natl Acad. Sci. USA* **90**, 3343–3347 (1993).
- Covello, P. & Gray, M. W. *Trends Genet.* **9**, 265–268 (1993).
- Schultz, D. W. & Yarus, M. J. *Mol. Biol.* **235**, 1377–1380 (1994).
- Janke, A., Feldmaier-Fuchs, G., Gemmell, N., von Haeseler, A. & Pääbo, S. *J. Mol. Evol.* **42**, 153–159 (1996).

Correction

In the reply by D. Tilman, C. Lehman, R. May and M. Nowak to S. Budiansky in Scientific Correspondence of 18 July 1996 (*Nature* **382**, 215–216; 1996), the third full paragraph, line two, the number that appeared as '104' should in fact have been ten raised to the fourth power, or 10,000. □

Effects of radiation on children

SIR — Dubrova *et al.*¹ reported that mutation rates at minisatellite loci in 79 children of parents who lived in heavily polluted areas of Belarus after the Chernobyl accident were twice that of 105 control children from the United Kingdom. They suggested that initial acute exposure to iodine-131 or chronic exposure of the parents to caesium-137 was responsible for the increased mutation rates, although they noted that the individual dose of ¹³⁷Cs “was estimated to be less than 5 mSv per year, a value far

exposed parent, therefore fewer than 12 mutations were derived from exposed gametes. Although the total number of bands determined for the exposed gametes in our study is estimated to be 900, approximately half of that in Belarus, we detected no increase of mutation rates at very similar minisatellite loci.

Dubrova *et al.*¹ did not match parental ages as well as genetic and environmental factors affecting background mutation rates of the two populations. Before it is possible to conclude that the observed

s.d. = 6.05, respectively ($t = 0.90$, $P > 0.05$; $\chi^2 = 2.58$, d.f. = 1, $P > 0.05$). The increase in mutation rate in the exposed group cannot therefore be explained by greater parental age.

We also discussed in our paper the possibility of mutation induction by other environmental pollutants. We stress that we found similar increases in mutation rate for three independent minisatellite systems. We have now extended these studies using additional loci, and again see an elevated mutation rate in the offspring of the same irradiated parents (unpublished data). We therefore believe that the difference in mutation rate found between the control and exposed families is most probably caused by environmental factors, including post-Chernobyl radioactive contamination. We also note that Satoh and Kodaira do not mention our dose-response analysis within the Belarus sample (Fig. 4 in our paper), which is consistent with the hypothesis of radiation induction of minisatellite mutation, and for which the control sample is irrelevant.

Satoh and Kodaira's own study on 5 minisatellites does not provide evidence for radiation-induced increase in mutation rate in the offspring of atomic bomb survivors in Hiroshima and Nagasaki. They also provide additional multilocus DNA fingerprint data which again point to no mutation increase. This apparent discrepancy could reflect the totally different nature of exposure in Japan following the atomic bomb explosions to that in the Chernobyl accident.

In summary, we note that our data, as well as those of Satoh and Kodaira⁵, are derived from relatively small numbers of families and that additional population surveys are needed to investigate the mechanisms and magnitude of minisatellite mutation induction seen in our study.

Yuri E. Dubrova*

Rita Newmann

David L. Neil

Alec J. Jeffreys

*Department of Genetics,
University of Leicester,
Leicester LE1 7RH, UK*

Valeri N. Nesterov

Nicolay G. Krouchinsky

Vladislav A. Ostapenko

*Research Institute for Radiation Medicine,
Mogilev Branch,
Mogilev 212004, Belarus*

1. Dubrova, Y. E. *et al.* *Nature* **380**, 683–686 (1996).

2. Neel, J. V. *et al.* *Proc. Natl Acad. Sci. USA* **93**, 2690–2695 (1996).

3. Sevan'kaev, A. V. *et al.* *Int. J. Radiat. Biol.* **63**, 361–367 (1993).

4. Scheid, W. *et al.* *Health Phys.* **64**, 531–534 (1993).

5. Kodaira, M. *et al.* *Am. J. Hum. Genet.* **57**, 1275–1283 (1995).

6. Jeffreys, A. J. *et al.* *Nature* **332**, 278–281 (1988).

7. Vogel, F. & Rathenberg, R. *Adv. Hum. Genet.* **5**, 223–318 (1975).

*Also at: N. I. Vavilov Institute of General Genetics, Russian Academy of Sciences, Moscow B-333, Russia. To whom correspondence should be addressed: e-mail: yed2@ie.ac.uk

MUTATIONS IN DNA FINGERPRINTS OF THE CHILDREN OF ATOMIC BOMB SURVIVORS

Group	No. of children	Total no. of bands* in children	Bands per child	No. of mutations	Mutation rate per band
Exposed	64	1,111	17.36	12	0.011
Control	60	1,041	17.35	13	0.012

A probe (15.1.11.4), which includes the core region from DNA fingerprint probe 33.15, was used to detect the fingerprints.

*DNA fragments larger than 2.3 kb were counted.

below that predicted from mouse and human data.” But they do not exclude the possibility of other contaminants such as industrial and agricultural pollutants, to which we add virus infections such as JC polyoma virus, which is suggested² by the existence of rogue cells with extreme chromosome damage observed in Belarus^{3,4}.

We recently reported a study⁵ of children of survivors of the Hiroshima/Nagasaki atom bombs, in which we reported no genetic effects at six minisatellite loci after examining 50 exposed families with 64 children and 50 control families with 60 children. Except for parental exposure to radiation, the genetic and environmental backgrounds are identical and mean parental ages at the birth of the children are indistinguishable in the two groups. The mean mutation rate per locus per gamete at the six loci was 1.5% in the exposed, whose estimated mean dose was 1.9 Sv, and 2.0% in the unexposed gametes. These values are similar to spontaneous germ-cell mutation rates detected in the Centre d'Etude du Polymorphisme Humain panel of 40 families⁶ and the Dubrova *et al.*¹ control group.

In preliminary examination of families in our study of the identical DNA fingerprints examined by Dubrova *et al.*¹, we again could not detect any effects of radiation (see table). Contrary to the bands detected with single-locus probes, fingerprint bands of children were sometimes difficult to trace back to one parent, and we scored 12 and 13 mutations, respectively, for the exposed and the control families. Except for one child, each child of an exposed family had only one

increase of the mutation rate in Belarus was caused by parental radiation exposure, it will be necessary to select new control Belarus families in which children born before the accident and parental ages at the birth of the children are matched to the parents having children after the accident, or at least looking at children born before the accident in the Belarus families already examined.

Chiyoko Satoh

Mieko Kodaira

*Department of Genetics
Radiation Effects Research Foundation
5–2 Hijiyama Park,
Minami-ku,
Hiroshima 732, Japan
e-mail: csatoh@rerf.or.jp*

DUBROVA *ET AL.* REPLY — We appreciate Satoh and Kodaira's point that the most appropriate control for our study would have been older children born to the same families before the Chernobyl accident. We discussed in our paper the difficulty of obtaining these controls, which forced us to use families from the United Kingdom as a control¹. We would also like to comment on additional points raised by Satoh and Kodaira.

Satoh and Kodaira suggested that we should have matched the parental age at birth, given that mutation rate is known to increase with parental age for at least some classes of mutation⁷. The maternal age for the control and exposed groups in our study was 24.41 ± 0.66 , s.d. = 5.35 and 24.53 ± 0.50 , s.d. = 5.58, respectively ($t = 0.14$, $P > 0.05$; Bartlett test for homogeneity of group variances, $\chi^2 = 0.14$, d.f. = 1, $P > 0.05$); the paternal age was 26.84 ± 0.88 , s.d. = 7.18 and 27.72 ± 0.55 ,

God in the details

Jerry A. Coyne

Darwin's Black Box: The Biochemical Challenge to Evolution. By Michael J. Behe. Free Press/Simon and Schuster: 1996. Pp. 307. \$25.

THE goal of creationists has always been to replace the teaching of evolution with the narrative given in the first eleven chapters of Genesis. When the courts stymied this effort, creationists tried a new strategy: cloaking themselves in the mantle of science. This produced the oxymoronic 'scientific creationism', arguing that the very facts of biology and geology show that the Earth is young, all species were created suddenly and simultaneously, and mass extinctions were caused by a great worldwide flood. The resemblance between this theory and the book of Genesis was, of course, purely coincidental. Scientific creationism, however, also came to grief. Virtually all creation 'scientists' were religious fundamentalists without biological expertise, and American courts clearly spied clerical collars beneath the lab coats.

In *Darwin's Black Box*, Michael Behe offers a new and more sophisticated version of scientific creationism. Unlike his predecessors, Behe is a genuine scientist, a biochemist from Lehigh University in Pennsylvania. The book jacket asserts that he is not a creationist, but believes in the scientific method. His argument, however, is a recycled version of the creationist notion that 'complex design' implies an intelligent designer. But where William Paley illustrated this logic with a watch, Behe uses biochemistry. His intended audience of lay readers may be impressed by the elaborate descriptions of molecular biology and long lists of references, but Behe's 'scientific' alternative to evolution ultimately becomes a confusing and untestable farrago of contradictory ideas.

Behe's thesis is that organisms harbour molecular pathways so elaborate and interconnected that they cannot be explained by gradual evolution from simpler precursors. His examples of such pathways, described with admirable clarity, include blood-clotting, the immune system and intracellular transport. These share what he calls "irreducible complexity": they would not function if any single component were removed. Because Darwinism requires that a pathway be useful at every stage of its evolution, Behe claims that such irreducibly complex pathways could not evolve in steps. Their existence therefore implies conscious design and an intelligent designer. (Like all scientific creationists, Behe keeps quiet about the identity of the Great Designer, but the author's professed Roman Catholicism offers one clue.) Evo-

lutionists are said to resist this idea of design because of our dogged but unreasonable dislike of supernatural explanations. Behe, however, is free from this constraint. With paternal pride, he declares that his discovery of biochemical design "must be ranked as one of the greatest achievements in the history of science", rivalling "those of Newton and Einstein, Lavoisier and Schrödinger, Pasteur, and Darwin".

There is no doubt that the pathways described by Behe are dauntingly complex, and their evolution will be hard to unravel. Unlike anatomical structures, the evolution of which can be traced with fossils, biochemical evolution must be reconstructed from highly evolved living organisms, and we may forever be unable to envisage the first proto-pathways. It is not valid, however, to assume that, because one man cannot imagine such pathways, they could not have existed. Moreover, as J. B. S. Haldane pointed out: "My own suspicion is that the universe is not only queerer than we suppose, but queerer than we *can* suppose." We face not only an absence of data, but also the awful fact that we ourselves are evolved creatures with limited cognition and imagination.

The answer to Behe's argument lies in realizing that biochemical pathways did not evolve by the sequential addition of steps to pathways that became functional only at the end. Instead, they have been rigged up

with pieces co-opted from other pathways, duplicated genes and early multifunctional enzymes. Thrombin, for example, is one of the key proteins in blood-clotting, but also acts in cell division, and is related to the digestive enzyme trypsin. Who knows which function came first? Behe makes a few half-hearted attempts to build up such pathways, but quickly abandons the enterprise and cries "design".

Evolutionists will find two other problems with Behe's arguments. First, there is ample evidence for the evolution of morphology and anatomy from studies of palaeontology, embryology, biogeography and vestigial organs. Such evolution must, of course, be based on the evolution of molecules and biochemical pathways. Second, we have plenty of direct evidence for the evolution of molecules. This includes the remarkable congruence between phylogenies based on anatomy and those based on DNA or protein sequence (bat haemoglobin, for example, is far more similar to that of whales than of birds), the relatedness of genes through gene duplication (including those involved in the immune system and blood-clotting), and the existence of vestigial 'pseudogenes' that were useful in ancestors. (Unlike most mammals, humans cannot synthesize vitamin C; we still carry the gene for the final step in this pathway, but deletions have rendered it non-functional.)

Behe's response to these problems constitutes the major weakness of his theory. He chews on the idea of morphological evolution, but cannot bring himself to swallow it. He finds the idea of common descent of all organisms "fairly convincing", and admits that microevolution occurs within species, but sees no evidence for transitions between major forms. (How one can admit common descent but deny

A FEMALE mosquito embedded in amber. Taken from *Amber: Window to the Past* by David A. Grimaldi. This book serves as the catalogue for an exhibition at the American Museum of Natural History in New York which includes flora and fauna preserved in amber as well as a wide range of amber jewellery and artefacts. Harry N. Abrams, \$49.50.



macroevolution is one of the fascinating questions Behe leaves unanswered.) Finally, in a tactic unique in the creationist literature, he admits that *both* evolution and creation might occur at the molecular level. Such a hybrid theory, however, yields sterile offspring, such as Behe's idea that the first 'designed' cell could include the DNA for all future evolutionary change, including that producing eyes and the immune system.

Responding to observations of non-functional genes and inefficient molecular processes, Behe theorizes that the Great Designer, like his counterparts in Paris and Milan, has goals beyond functionality: "Features that strike us as odd in a design might have been placed there by the designer for a reason — for artistic reasons, for variety, to show off, for some as-yet-undetected practical purpose, or for some unguessable reason — or they might not." One should add the "puckish reason": to confuse future biologists by making things look as though they evolved.

If one accepts Behe's idea that both evolution and creation can operate together, and that the Designer's goals are unfathomable, then one confronts an airtight theory that can't be proved wrong. I can imagine evidence that would falsify evolution (a hominid fossil in the Precambrian would do nicely), but none that could falsify Behe's composite theory. Even if, after immense effort, we are able to understand the evolution of a complex biochemical pathway, Behe could simply claim that evidence for design resides in the other unexplained pathways. Because we will never explain everything, there will always be evidence for design. This regressive *ad hoc* creationism may seem clever, but it is certainly not science. As the theologian Dietrich Bonhoeffer pointed out, it is also bad religion: "If in fact the frontiers of knowledge are being pushed farther and farther back (and that is bound to be the case), then God is being pushed back with them, and is therefore continually in retreat."

In the end, *Darwin's Black Box* is a work of advocacy whose creationist ancestry is revealed by both its rhetoric and its failure to deal honestly with the evidence for evolution. There is the usual selective quotation of evolutionists (including, to my horror, a remark of my own, both altered and taken out of context), ridicule of scientists, and a folksy 'us-against-them' style reflecting the populist roots of creationism. The book will no doubt be widely cited by Biblical creationists, who will tout its message of design while ignoring its timid acceptance of evolution and its view of the creator as Cosmic Prankster.

If the history of science shows us anything, it is that we get nowhere by labelling our ignorance 'God'. Lord Kelvin declared that the *primaeva* Earth had cooled down too quickly to permit the great age required by Darwinism. How could he

imagine that radioactivity would be discovered four decades later and prove the missing source of heat? Duane Gish, the doyen of American creationists, once argued for the separate creation of mammals and reptiles, based on their jaw structure. Each has a jaw joint made from a different pair of bones, and Gish could not imagine how the transitional form could chew while its jaw was being unhinged and rearticulated. In 1958, however, Fuzz Crompton described a mammal-like reptile with a double jaw joint that included both pairs of bones. The evolution of biochemical pathways is certainly queerer than Professor Behe can suppose. □

Jerry Coyne is in the Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA.

The end of the beginning

John D. Barrow

Time's Arrow and Archimedes' Point: New Directions for the Physics of Time. By Huw Price. Oxford University Press: 1996. Pp. 306. \$25, £18.95.

"TIME flies like an arrow; fruit flies like a banana", according to the Marx Brothers' analysis of the vexed question of the arrow of time. This temporal conspiracy at the heart of things has many subplots: why do we have memories of the past, why is the Universe expanding rather than contracting, why do measures of disorder grow with time in closed systems, while the equations of physics seem to possess an entirely democratic attitude with respect to the direction of time?

This book is an attempt to grapple with the different problems of the arrow of time with a high degree of analytical care. Separate chapters deal with CP-violating systems in elementary particle physics, the reduction of the wave function and the problem of time in quantum mechanics, and advanced and retarded radiation, but there is a special focus of interest on cosmology as the crux of the arrow-of-time problem. For the most part, the treatment is wide-ranging and substantial. The discussion is carefully signposted and chapters contain point-by-point summaries of the argument and the author's principal conclusions.

The author announces at the start that he adopts the "block universe" view of spacetime as an objective entity, as opposed to imagining that time is just an artefact of a special human perspective on cosmic events. This is the backdrop for his atemporal or 'Archimedean' view of time in which the past and future are dealt with

even-handedly. He is keen to uncover errors of reasoning that have infested many recent analyses of the origins of the arrow of time in physical problems, for example attempts to distinguish initial and final states of closed universes using quantum cosmological boundary conditions, attempts to do the same using inflation, and fallacies in the use of the Wheeler-Feynman absorber theory. In these aims he succeeds with great clarity. His arguments are accessible to physicist and philosopher alike. However, moving to the more general cosmological claims, there should be surprise at the way in which some cosmological speculations are adopted as fact and used as the basis for the discussion of the cosmological time-arrow problem.

The author makes great play of his analysis of cosmological arrows of time, in particular the often-asked question of why the entropy of the Universe increases. He claims that this problem is misconceived in some sense: that the real problem is not to explain the trend but "to explain the low-entropy past" of the Universe. The discussion of cosmology revolves around this issue. The problem with his approach is twofold: first, we do not have any reliable measure of 'the entropy of the Universe' (nor do we even know that such a quantity exists); second, we do not know that the Universe began in a low-entropy state, as the author assumes. In fact, one might claim that the situation is even more fundamentally uncertain: there is no way we can ever know whether or not the Universe began in a low-entropy state.

The problem of specifying the entropy of the Universe is that we do not know all the potential contributors to it. There are classical and quantum contributions, dynamical contributions, numbers of particle species, and so on. But there is no reason to think that the entropy of even the visible part of the Universe is fully measured by counting the massive and massless particles it contains, or simply by calculating something like the Bekenstein-Hawking entropy of a black hole the size of the particle horizon. Neither of these measures keeps track of all the forms of disorder that can distinguish one universe from another. They do not have the properties we would expect of measures of cosmological disorder and information. To obtain a measure of the visible universe's entropy, we need to solve the great problems posed by the study of generic forms of complexity: what is complexity and how do we quantify it in a way that tells us whether one thing is more complex than another? There are many candidates, each satisfactory for some types of complexity, but none appropriate for all.

More problematic than the question of a measure of the entropy of the Universe is the assumption that the entropy of the Universe was low in the distant past (at 'the

beginning' perhaps?). It is not clear where the author has taken this key datum from. I know of no sure evidence for it, even if we pick on simple intuitive measures of cosmological non-uniformity to measure its entropy. Nor could there be any such evidence. Our observations of the Universe allow us to sample its structure only on and close to the inside of our past light cone. Observations of the microwave background radiation isotropy and deductions from primordial nucleosynthesis allow us to go back some way in time but still leave us very far from establishing the character of the beginning of the expansion. Even with perfect observations, we could learn only about the structure of a minuscule part of the structure of the Universe at a moment of cosmic time close to the apparent beginning. All the information we would need to establish whether the whole initial state was of low entropy is not causally accessible to us even with perfect observations (neutrino telescopes, graviton telescopes) because of the finite speed of light. Price seems to be able to overcome this limitation only by assuming what is to be proved.

The inflationary universe theory, which is consistent with current observations and should be partly tested by satellite-borne instruments over the next decade, opens up a Pandora's box of further possibilities and probabilities that make it even harder to defend the assumption that the Universe began in a state of very low entropy. Inflationary universes can begin in highly irregular states of high entropy, far from

equilibrium, and end up appearing very regular to observers such as ourselves. Different regions can undergo different amounts of inflation (which may even continue for all future time like a stochastic branching process). Our visible Universe (a finite part of the possibly infinite, highly inhomogeneous whole) is all that we can see. The entropy of the whole Universe is not accessible to us. Indeed, we might expect to inhabit a fluctuation with special features that permit expansion to last long enough to create the conditions in which the evolution of complexity is possible. Inflation predicts just that, but we can never test that prediction.

For these reasons, despite the fact that the author has done physicists a great service in laying out so clearly and critically the nature of the various time-asymmetry problems of physics, I have reservations about the author's discussion of the cosmological arrow of time. It is really only a discussion of particular speculative assumptions about cosmological initial conditions. In grappling with the symmetries and asymmetries of cosmological time, we still seem to be nearer the end of the beginning than the beginning of the end. □

John D. Barrow is in the Astronomy Centre, University of Sussex, Brighton BN1 9QH, UK.

■ *The Anthropic Cosmological Principle* by John D. Barrow and Frank J. Tipler has just been published by Oxford University Press in a new paperback edition, price £9.99.

Origins of clock watching

Owen Gingerich

History of the Hour: Clocks and Modern Temporal Orders. By Gerhard Dohrn-van Rossum. *University of Chicago Press*: 1996. Pp. 455. \$29.95, £23.95.

APART from flights, telephones and television, one of the most significant ways in which life in the twentieth century differs from antiquity is the pervasive presence of clocks. Our lives are scheduled around specific times and intervals in a fashion that would have been impossible in ancient Rome.

When and how did this profound change come about? Did monasteries acquire the first clocks and then, in league with an emerging merchant class, enforce their own rigid time standards on the surrounding populace? This view — the accepted interpretation for at least 50 years — is carefully dissected and challenged in Dohrn-van Rossum's revisionist account.

The author, a specialist in mediaeval and early modern history at the Univer-

sity of Bielefeld in Germany, took up this investigation when a television network asked for help in finding images to illustrate "church's time" and "merchant's time". This led to a "gigantic puzzle" to assemble information from 1,000 European cities and towns, and to the discovery that the standard picture was seriously defective.

In a nutshell, Dohrn-van Rossum argues that the striking clock was invented in Europe, perhaps independently several times, around 1280. The sound of bells across Europe was not new, but the method of regulation was, and by the early 1300s striking clocks became objects of prestige that helped to put towns on the map. The presence of audible timekeepers, followed by visible ones, gave a new sense of time that began to permeate the workplace, whether this involved classes at schools, weavers, merchants, the harvesters of grapes, or goldsmiths.

But merchants as a class and the Church were not especially at the fore-

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

A household admiring the master's rare clock, from a mid-fifteenth-century Flemish treatise on morality.

front of this radical and pervasive change. Rather, the driving force was the desire of many different groups to acquire the latest status symbol for their communities.

This is the main thrust of the book. The author has produced an extraordinarily detailed, and at times almost tedious, *tour de force* of research. His use of pictorial sources is impressive, especially when they fill in pieces of the puzzle missing in the texts. On a subject as vast and intricate as this, no one will ever have the final word, but at least no future explorer of time can afford to ignore this remarkably comprehensive study.

This rich lode of material can provide a perceptive background for problems the author touches only peripherally. For example, when in the early 1600s Galileo began to quantify time in his discussion of acceleration and free fall, he was working with a newly formed numerative consciousness of time that Aristotle could not have experienced.

Despite the wide range of sources the author has brought to bear on the changing role of time in the late Middle Ages, he has not ploughed all the ground. Nicholas Copernicus, we are told, was born at 4.48 p.m. on 19 February 1473. In the absence of a baptismal record, we would not even know the day of his birth if we did not also have the hour and minute, preserved in a Renaissance horoscope. It is unlikely that the physician or midwife had a clock at the bedside. The time was probably calculated later by one of several methods used by astrologers. The desire for accurate horoscopes seems to have marched in parallel with the arrival of public clocks. Dohrn-van Rossum rather dismisses the influence of astrologers and astronomers in creating a demand for, and a sense of the significance of, time, but I suspect that this is a fertile field for additional research.

Should this important book go into a second edition, the author will have a chance to make a few small corrections. In a book entitled *History of the Hour*, it is embarrassing to have the origin of the hour erroneously placed in Babylon instead of in the very different astronomical culture of ancient Egypt. The translation from the German reads quite well, but in a book that often uses both 'canon' and 'cannon', no doubt the frequent misspelling of 'cannon' arises from the similarity in German of *Kanon* (canon) and *Kanone* (cannon).

There is a considerable muddle in the final chapter where the German *Meile* is repeatedly translated as "mile" instead of "league". In one instance the correct reading of 764 kilometres in the German has been altered so that the English sentence, describing the speed of post riders, reads: "One hundred and thirty-two hours were needed for 103 miles (164 kilo-metres)."

These quibbles aside, Dohrn-van Rossum has produced a persuasive and brilliantly documented new understanding of how modern time-consciousness arose.

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

The need to love technology

Wiebe E. Bijker

Aramis, or the Love of Technology. By Bruno Latour. Harvard University Press: 1996. Pp. 314. \$45, £29.95 (hbk); \$1995, £13.50 (pbk).

ARAMIS was to have been a fully automated, personal rapid transit system for Paris, combining the advantages of the car (individual, comfortable, point-to-point transportation without changing vehicles) with those of the train (no traffic jams and less pollution, yet accessible to everyone). But it is no more. So who killed it?

The first vehicles ran on a test track in 1987, and a review commission concluded that the system "could be brought to a level of technological realization" and that there was a market for it. But the project was cancelled abruptly, at a cost of 17 years of research and millions of francs. There was no scandal in the press, no bitter accusations between the contract partners and no political heads rolled. Nevertheless, all parties come under suspicion in this whodunnit — the French company Matra, because it was perhaps more interested in investing its resources

in the Lille metro; the Matra engineers, because the project was never really technically feasible; the Parisian public transport agency RATP, because the vehicles had seats but no standing room, so passengers could not be squeezed in during rush hours; the budget office of the French ministry of finance, because Aramis was too expensive; and the intended passengers, because they were frightened to lose the anonymity of the Metro.

Immediately after the project ended, Bruno Latour was asked by the RATP to investigate what went wrong. On the basis of a detailed empirical study, he has written three books in one: a detective novel, in which a sociology professor and a young engineer play the parts of Sherlock Holmes and Dr Watson; a scholarly treatise introducing the modern sociology of technology; and a reproduction of original archival documents. As the book develops, we hear the voice of technology itself, with Frankenstein's "humachine" and Aramis himself as spokespersons.

The young engineer's initiation into the sociology of technology begins with the demise of all of his standard explanations for the failure of Aramis — that it was technically unfeasible, that there was no market for it, that it was too expensive, and that it had insufficient political backing. On each and every point, his professor produces documents and interview excerpts that show exactly the opposite. Even more important is another act of debunking: technology is not something cold, nonhuman and antisocial. Technology is to be "loved", for its own sake and as a basic constituent of modern society and human affairs. (Here 'love' does not mean 'adore', or even 'evaluate positively', but rather 'engage in'.) This serves as a prelude to Latour's final conclusion — that an isolated technology is doomed.

In the sociological treatise, this line of analysis is characterized as symmetrical and relativist. It is symmetrical in that successful and failing outcomes are treated in the same manner; only by such a symmetrical analysis can we hope to gain an understanding that does more than "flatter the victors of the day" and provide a comfortable conclusion "at the end of the road, when we've settled down by the fire". And it is relativist because it assumes that nothing exists independently of what the actors do; that there are no fixed frames of reference; "actors never swim twice in the same river". So the economics of the project and the country, the consumer demand, the technology's intrinsic safety and the project's political support all need to be constructed continuously by the actors.

The young engineer and his professor discuss the details of the technology when investigating the last phase of the project. In a passage as hilarious as it is perceptive,

the two investigators take on the roles of the software components of Aramis. This allows the engineer to get on top of the matter, and observe that his supervisor "lost his grip and could no longer write any 'sociological commentary,' as he pompously called it, on what we were discovering". The professor suffers a nervous breakdown and the student must finish the report himself.

This breakdown is symbolic of the role that Latour denies to theoretical explanations and generalization beyond one case study: "I'm not looking for anything else [than] a refined sociology which applies to a single case, to Aramis and only Aramis. A single explanation, for a single, unique case; then we'll trash it", his professor had remarked long before the breakdown. Detailed accounts of all the network building, the negotiations, the constructive work by engineers, politicians, managers and passengers are all that his relationist sociology offers, he claims: "there are as many theories of action as there are actors". This is a case of feigned modesty, probably for pedagogical reasons to chastise classical sociology.

Latour's book does offer important insights into the sociotechnical domain and engineering practices that transcend the Aramis case. It also provides, mainly in the form of methodological discussions, the groundwork for a theory of technology and society. This important asset, of what I think is Latour's best book so far, is not diminished by his obvious difficulty in reconciling his aversion to sweeping reductionist concepts with his urge to contribute to a general understanding of technology. The next challenge is to develop such a theory without falling into the trap of *a priori* fixed categories and losing the newly won insights from detailed analysis of technological projects.

So who did kill Aramis? After studying one of the last documents of October 1987, our young engineer suddenly finds the solution and can deliver his masterpiece: "The report presented the 1987 Aramis, word for word, as identical to [the] 1970 Aramis. I myself had found twenty-one interpretations, but the technological documents remained mute about this dispersion. Aramis had not incorporated any of the transformations of its environment. It had remained purely an object. Remote from the social arena, remote from history." Aramis had, in the end, been left intact, untouched, unmodified, unloved. Nobody had loved it enough to make it grow and develop. All parties were guilty—as a true reflection of another train murder, on the Orient Express.

Wiebe E. Bijker is in the Department of Technology and Society Studies, Rijksuniversiteit Limburg, 6200 MD Maastricht, The Netherlands.

Atmospheric gas concentrations over the past century measured in air from firn at the South Pole

M. Battle*, M. Bender*, T. Sowers†, P. P. Tans‡, J. H. Butler‡, J. W. Elkins‡, J. T. Ellis*, T. Conway‡, N. Zhang‡§, P. Lang‡ & A. D. Clarke‡§

* Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882, USA

† Geosciences Department, Penn State University, University Park, Pennsylvania 16802, USA

‡ National Oceanic and Atmospheric Administration, Climate Monitoring and Diagnostics Laboratory, Boulder, Colorado 80303, USA

§ Cooperative Institute for Research in Environmental Studies, University of Colorado, Boulder, Colorado 80309, USA

The extraction and analysis of air from the snowpack (firn) at the South Pole provides atmospheric concentration histories of biogenic greenhouse gases since the beginning of the present century which confirm and expand on those derived from studies of air trapped in ice cores. Furthermore, calculations based on the inferred atmospheric concentrations of oxygen and carbon dioxide indicate that—in contrast to the past few years—the terrestrial biosphere was neither a source nor sink of CO₂ between ~1977 and 1985.

OUR knowledge of the human effects on the content of the atmosphere comes from direct measurements and, for earlier times, studies of the composition of air trapped in polar ice cores. Although the ice-core records are extremely informative, they have limitations. The limited size of ice cores constrains the amount of occluded air available for analysis. This restriction limits the precision with which some chemical species can be measured. Furthermore, the atmospheric composition reconstructed from ice cores may be compromised by artefacts associated with extraction of the occluded air, or *in situ* chemical reactions between the ice and the air¹.

The firn (the upper layer of unconsolidated snow) of an ice sheet is porous and permeable. Consequently, interstitial air at a given depth exchanges with the air above and below by molecular diffusion. Changes in the composition of the atmosphere propagate downwards through the firn, such that the mean age of the air in the firn increases monotonically with depth. As first demonstrated by Schwander *et al.*², samples of firn air can be drawn from discrete depths, and pumped into flasks at the surface for future analysis. By collecting litres of air, one eliminates many artefacts and analytical limitations associated with extracting small amounts of gas from ice. Consequently, we are able to achieve analytical precision close to that of direct atmospheric samples.

We have analysed the firn air for CO₂, CH₄, N₂O, δ¹⁵N and O₂/N₂, and interpret the data using a simple model in which gas concentrations in the firn change owing to molecular diffusion. The maximum age of our samples is limited to roughly 100 years by the diffusive properties of the firn and the formation of closed bubbles in the ice. We show that the covariation of CO₂ and CH₄ in South Pole firn air is consistent with predictions of the diffusion model and the recent atmospheric concentration histories of CO₂ and CH₄ inferred from ice cores and measured by direct atmospheric sampling. This consistency shows that the inferred concentration histories are robust. Furthermore, the covariation of atmospheric N₂O and CO₂ shows a surprising linearity in the years since 1900.

The covariation of O₂ and CO₂ in South Pole firn air indicates that the land biosphere was neither a source nor a sink of atmospheric CO₂ during the years 1977–85, implying that carbon losses from deforestation were approximately balanced by net CO₂ uptake elsewhere. This result is consistent with model-derived

estimates of terrestrial carbon uptake for the same period^{4,5}. This balanced biosphere stands in contrast to recent measurements of terrestrial carbon uptake which show a substantial land sink in the years between 1992 and 1994^{6–8}. Unfortunately, artefacts associated with the sealing of air bubbles in ice prohibit inference of the carbon balance during earlier times.

Influences on the composition of firn air

At least five important processes influence the composition of firn air. First, the upper ~10 m can be mixed convectively by surface wind stress⁹. Second, the composition of the gas mixture at a particular depth reflects the chemical potentials of each component species. For each species, as for the total pressure, the equilibrium concentration increases with depth according to the barometric equation^{3,10–12}:

$$P_z/P_0 = \exp(mgz/RT)$$

where P is pressure, z is depth ($z = 0$ at the ice-sheet surface), m is molar mass, g is gravitational acceleration, R is the ideal gas constant, and T is absolute temperature. Consequently, heavier gases and isotopes become progressively enriched with depth.

Third, molecular diffusion causes changes in the composition of the overlying atmosphere to be reflected in the composition of the interstitial firn air. In the absence of gravity, these changes are described by Fick's second law:

$$\frac{\partial C}{\partial t} = \frac{\partial}{\partial z} \left(D(z) \frac{\partial C}{\partial z} \right)$$

where C is the concentration of a particular species, t is time and D is diffusivity. In the firn, open porosity falls with depth, and tortuosity rises, causing diffusivity to decrease^{3,13}. The diffusive communication between air masses throughout the firn column means that air at any given depth reflects a weighted mean of atmospheric composition over some interval of time.

Fourth, some O₂ appears to be excluded from the air bubbles trapped in ice^{11,12,14}. The excluded O₂ diffuses up through the firn and escapes to the atmosphere, altering the composition of the firn air along the way. Exclusion of O₂ apparently takes place during bubble close-off, when channels of molecular size form, allowing the relatively prolate O₂ molecules to escape more

readily than N_2 (ref. 11). This exclusion is probably not limited to O_2 . However, in most other species the large size of anthropogenic changes in the atmosphere masks the effects of preferential exclusion.

Fifth, thermal gradients in the firn cause an enrichment of heavier species and isotopes in its colder regions^{15–17}. The enrichment occurs because diffusion of heat in firn is an order of magnitude slower than molecular diffusion of gases^{2,18}. Thermal gradients arise from both long-term (decadal) warming¹⁹ and seasonal thermal cycles. Long-term warming causes the top of the firn to be warmer than the bottom, so heavier gases and isotopes will be progressively enriched with depth. For the work presented here, long-term gradients are considered negligible. Seasonal temperature changes lead to substantial thermal and

concentration gradients which reverse semi-annually. Among the species considered here, seasonal thermal effects are significant for $\delta^{15}N$ and O_2/N_2 only. Below 50 m, seasonal thermal gradients are vanishingly small, and the thermal inertia of the firn is such that global warming of the past 40 years is unimportant. Thus, for O_2/N_2 we consider only samples taken from below 50 m when interpreting our results.

Sampling and observations

Our goal in this study was to characterize the composition of old firn air, thereby constraining anthropogenic changes in atmospheric chemistry that predate direct measurements. To this end, we chose to collect samples of firn air from two boreholes at the South Pole (90° S). It is an extremely cold site (annual mean temperature $-49.4^\circ C$) with a snow accumulation rate of ~ 8 cm ice-equivalent per year²⁰. This leads to the deepest known firn-ice transition (123 m; refs 13, 21), and in turn, unusually old air at the base of the diffusive column.

Samples were analysed for CO_2 , CH_4 , N_2O , $\delta^{15}N$ of N_2 and $\delta O_2/N_2$ (The last two quantities are defined in Fig. 1 legend). We use the measured $\delta^{15}N$ to correct for the effects of gravitational fractionation on the concentrations of other species^{12,22}. Values of $\delta^{15}N$ and $\delta O_2/N_2$ are reported in 'per meg' where 1 per meg $\equiv 0.001\%$. For CO_2 , CH_4 and N_2O , the agreement of replicate flasks (see Fig. 1 legend) implies very little sampling error for these gases. However, the first flasks of sample pairs collected between 20 and 100 m depth are consistently elevated in $\delta^{15}N$ by 4 ± 1 ($n = 14$) per meg, and in $\delta O_2/N_2$ by 22 ± 2 per meg (11 ± 3 after ^{15}N correction). This bias is at present unexplained. If it were due to contamination with shallower firn air in the second bottle, we would expect CO_2 to be elevated by $0.18 \mu mol mol^{-1}$ in the second flask. However, we observe no significant CO_2 difference ($0.007 \pm 0.03 \mu mol mol^{-1}$).

The O_2/N_2 bias is problematical because it might point to aliasing of our O_2/N_2 -depth profile. Specifically, $\delta O_2/N_2$ in our firn samples may differ from the true firn values by some depth-dependent amount. There are three lines of evidence against such aliasing. First, $\delta O_2/N_2$ of surface air collected with our apparatus is offset from independent measurements (R. Keeling, personal communication) by <8 per meg. Second, the offset in $\delta O_2/N_2$ between the first and second flasks is constant (both before and after correction for gravitational effects) over the critical depth interval of 60–95 m. Third, as discussed below, the O_2/N_2 time history inferred from the firn air data projects a modern value that is compatible with the present atmospheric ratio. Based on these observations, aliasing is highly unlikely, but it remains a remote possibility.

Figure 1 shows the observed depth dependence of $\delta^{15}N$ of N_2 , as well as gravitationally corrected profiles of CO_2 , CH_4 , N_2O and $\delta O_2/N_2$. As the $\delta^{15}N$ of tropospheric N_2 is constant, gravity and temperature are the only factors causing $\delta^{15}N$ to vary in the firn. Their effects are reflected in the $\delta^{15}N$ profile. The anomalous $\delta^{15}N$ enrichments at 10 and 20 m depth are caused by the seasonal thermal gradient (the same samples also show a thermally induced enrichment in $\delta O_2/N_2$). Between 50 and 115 m depth, $\delta^{15}N$ increases with depth at a rate of $5.12 \pm .05$ per meg per m, which agrees fairly well with the value of 5.30 per meg per m predicted by the barometric equation. Below 115 m depth, $\delta^{15}N$ is roughly constant. The absence of continued gravitational enrichment in the heavy isotope implies a 7-m-thick 'lock-in zone'. In this depth zone, impermeable winter layers prevent vertical diffusive mixing²³, but samples can still be extracted from summer layers containing some open porosity. Lock-in zones have also been observed at other sites^{2,24}.

The gravitationally corrected profiles of CO_2 , CH_4 and N_2O (Fig. 1) reflect the anthropogenic increases in the atmospheric concentrations of these gases over time. Deeper samples have lower concentrations, reflecting older mean ages. Concentrations decrease most rapidly within the lock-in zone near the bottom of the firn. Here, in the absence of vertical diffusive transport, the age

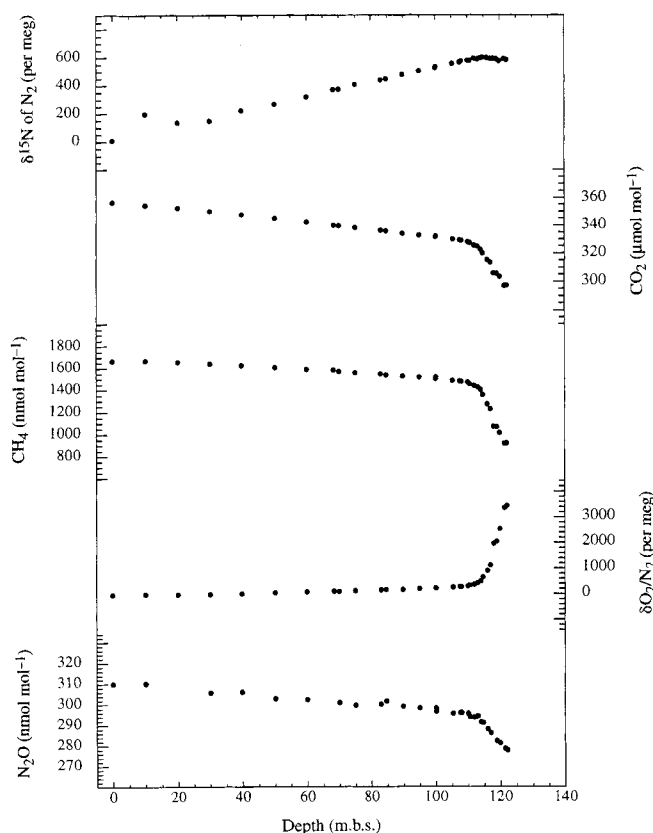


FIG. 1 Measured concentrations of gases as a function of depth (m.b.s., metres below surface). Values for CO_2 , CH_4 , $\delta O_2/N_2$ and N_2O have been corrected for gravitational fractionation. Values for $\delta^{15}N$ and $\delta O_2/N_2$ are given in 'per meg' defined as $\delta^{15}N = \{[(^{15}N/^{14}N)_{\text{sample}} / (^{15}N/^{14}N)_{\text{standard}}] - 1\} \times 10^6$ and $\delta O_2/N_2 = \{[(O_2/N_2)_{\text{sample}} / (O_2/N_2)_{\text{standard}}] - 1\} \times 10^6$. 1 per meg $\equiv 0.001\%$. The standard (GF1; ref. 7) is contemporary tropospheric air. Samples were collected from two 125-m holes, 30 m apart, 1.5 km upwind of Amundsen-Scott South Pole station. Firn air samples were collected at ~ 15 different depths per hole using a procedure similar to that of Schwander *et al.*². Samples of surface air were also collected, and core densities were measured at the drill site. Two samples were collected at each depth in Teflon-sealed glass flasks and analysed at the NOAA/CMDL laboratories for CO_2 (standard deviation $\sigma = \pm 0.06 \mu mol mol^{-1}$), CH_4 ($\pm 4.3 nmol mol^{-1}$) and N_2O ($\pm 1.1 nmol mol^{-1}$) (refs 48, 49). Two other samples were collected at each depth in Viton-sealed glass flasks and analysed at the University of Rhode Island for $\delta^{15}N$ of N_2 ($\sigma = \pm 2.6$ per meg) and $\delta O_2/N_2$ (± 5.4 per meg) (ref. 50). Note that we actually measure the mass ratio 32/29 for O_2/N_2 . Unless otherwise indicated, uncertainties presented refer to the 1σ reproducibility of measurements of a single flask. These errors reflect both instrumental precision and small artefacts associated with collection of the samples. They are calculated using the r.m.s. deviation from the mean of flask pairs collected at each depth (multiplied by $\sqrt{2}$).

of the air and the surrounding firn increase at the same rate. From 116.1 m to 121.3 m depth, CO_2 decreases from $315.3 \mu\text{mol mol}^{-1}$ (1959 air) to $297 \mu\text{mol mol}^{-1}$ (1903 air) (ref. 25). The 55-year interval implied by this decrease is roughly the same as the time required to accumulate 5.2 m of dense firn at South Pole, and confirms the lock-in zone inferred from the $\delta^{15}\text{N}$ record.

If the atmospheric concentration histories of CO_2 and CH_4 inferred from ice cores are accurate, they should account for the variation of CO_2 and CH_4 in the firn air. To test these inferred records, we follow Schwander *et al.*²⁹ and model transport in the firn with a finite-difference approximation of Fick's second law (given above) in one dimension. The model (Box 1) requires a diffusivity–depth profile to generate a concentration–depth profile for each species. This diffusivity profile is not well known. However, the modelled covariation of two species depends almost entirely on their atmospheric histories and the ratio of their diffusivities. For example, keeping the ratio of diffusivities constant while altering the absolute diffusivity for both species by $\pm 20\%$ (a full range of 40%) changes the model CH_4 prediction by less than 20 nmol mol^{-1} at any given CO_2 concentration. Thus, our test of the CO_2 and CH_4 histories is insensitive to the diffusivity–depth profile we assume for the model. Because we drive the

model with histories derived from the same ice core (DE08), relative dating uncertainties are not a source of error.

Figure 2 shows the observed covariation of CO_2 and CH_4 concentrations in the firn air, together with the prediction of the model described above. Agreement is generally excellent. A small discrepancy between data and model occurs for CH_4 concentrations below $\sim 1,300 \text{ nmol mol}^{-1}$. It is possible that very slow diffusive transport in the upper part of the lock-in zone might be responsible for this difference. This small anomaly notwithstanding, the agreement between the firn air data and the model predictions verifies the history of atmospheric CO_2 and CH_4 concentrations during the twentieth century inferred from ice-core data (ref. 25 and D. Etheridge, personal communication). Because of the deep diffusive column in the South Pole firn, atmospheric histories are highly convolved. Consequently, we are capable of resolving only the broad features of the ice-core record.

Atmospheric O_2 and inferred carbon fluxes

As described in detail elsewhere^{26–28}, the rise and fall of atmospheric O_2 directly reflects fossil-fuel combustion and terrestrial storage and release of carbon. If atmospheric O_2 falls faster (slower) than it is consumed by fossil-fuel burning, the cause must be a terrestrial O_2 sink (source), implying a release (uptake) of CO_2 by the land biosphere. The uptake or release rate of CO_2 by the land biosphere and the atmospheric history of CO_2 immediately gives the oceanic component by difference. In this approach, we assume that, over the decadal timescale of interest here, the seasonal flux of O_2 from the ocean to the atmosphere due to net primary production is in balance with the return flux to the oceans due to ventilation of O_2 -undersaturated waters⁷.

Records of atmospheric O_2/N_2 with the requisite precision extend back only to late 1988 at La Jolla, California²⁸, and for shorter periods at other sites^{7,8}. South Pole firn air thus allows us to extend the measurements of O_2/N_2 backwards by more than 10 years before the advent of direct atmospheric O_2/N_2 measurements.

Figure 1 shows that two major features dominate the O_2/N_2 profile below 40 m, where seasonal effects become negligible. First, between 50 m and 95 m there is a gradual increase in $\delta\text{O}_2/\text{N}_2$ with depth (and age) which is qualitatively consistent with the effusive consumption of O_2 and the production of CO_2 in combustion of fossil fuels. Second, below 95 m there is a dramatic rise in $\delta\text{O}_2/\text{N}_2$. As discussed above, this rise appears to result from selective effusion of O_2 out of closing bubbles and into firn air^{3,10–12}. The effusive O_2 input to the firn is most evident below 95 m, but it drives an upward flux of O_2 that must elevate the observed O_2/N_2 ratio all the way to the surface.

We use the firn air diffusion model described above to find a history of land-biosphere carbon fluxes that best allows us to reproduce the observed covariation of $\delta\text{O}_2/\text{N}_2$ and CO_2 in the firn. To do so, we model the O_2 –depth covariations linked to three processes: (1) fossil-fuel combustion (known), (2) land biosphere CO_2 uptake and O_2 release (unknown), and (3) O_2 exclusion from the firn during bubble close-off (unknown). We infer the uptake of CO_2 by the land biosphere and the O_2 -exclusion flux by finding the values of these two unknown terms that optimize the agreement between observed O_2 concentrations, and those predicted by the model. We consider only data collected below 50 m and above 114 m. Above 50 m, O_2/N_2 varies significantly due to the seasonal thermal cycle. Below 114 m, vertical diffusion of gases is impeded by lock-in. When fitting, we also allow the overall O_2/N_2 scale to float, in order to accommodate any depth-independent offset introduced while sampling the firn.

The excellent fit (Fig. 2) implies a biospheric source of $0.4 \pm 0.4 \text{ Gt C yr}^{-1}$ and an anomaly in the O_2/N_2 ratio of gas trapped in South Pole ice of $-8.5 \pm 0.6\%$ driving the up-hole O_2 flux. The O_2 anomaly implied by the size of the up-hole flux agrees with the value of $-8.1 \pm 0.4\%$ ($n = 6$) measured in samples of ice taken below the close-off in our first hole (corrected for gravitation and O_2 decrease since ~ 1900 due to fossil-fuel

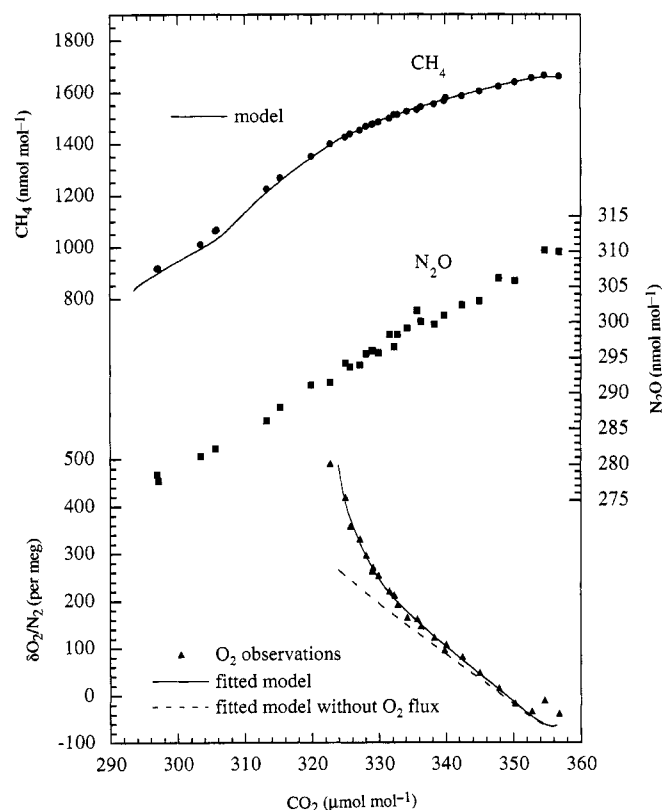


FIG. 2 Observations (gravitationally corrected; data points) and model output (lines) of the covariation with CO_2 of CH_4 , N_2O and $\delta\text{O}_2/\text{N}_2$ in the firn. The CH_4 – CO_2 (upper) plot shows model output when driven by atmospheric histories inferred from ice cores and direct sampling (see Box 1). The $\delta\text{O}_2/\text{N}_2$ – CO_2 (lower) plot shows a fit to the data that implies a biospheric source of $0.4 \pm 0.4 \text{ Gt C yr}^{-1}$ and an O_2/N_2 anomaly in the ice of $-8.5 \pm 0.6\%$ driving the up-hole O_2 flux. The O_2 history is derived from the fossil-fuel combustion records of Marland *et al.*⁵¹, the combustion stoichiometry of Keeling²⁷ and observed atmospheric $\delta\text{O}_2/\text{N}_2$ values after 1988^{7,28}. Errors in measured $\delta\text{O}_2/\text{N}_2$ values (average of two flasks at each depth) are ± 5.4 per meg, while errors in CO_2 are taken to be negligible. The optimized fit gives $\chi^2 = 34.6$ for 17 – 3 degrees of freedom for data taken from 60 to 114 m depth. The apparently anomalous observation at 10 m ($\text{CO}_2 = 354.4 \mu\text{mol mol}^{-1}$) is due to the seasonal thermal cycle.

burning). This similarity is reassuring, although the nearly exact agreement may be fortuitous as O_2/N_2 is somewhat variable in ice-core samples¹⁴.

Furthermore, the modelled O_2/N_2 extrapolated forward to the time of sample collection is only 22 per meg less than the South Pole surface samples, and ~16 per meg less than values based on the Cape Grim climatological O_2/N_2 record⁷. This small offset implies the near absence of aliasing in the samples from 60–113 m despite the difference between replicate flasks discussed above.

The uncertainties in the values given thus far reflect only the correlated errors associated with the fit. At least four other systematic errors need to be considered. First, uncertainty in the rates and stoichiometry of fossil-fuel combustion introduce an error of $\pm 10\%$ to the consumption rate of O_2 (or $\pm 0.8 \text{ Gt C yr}^{-1}$) (ref. 28). Second, the relative diffusivities of CO_2 and O_2 carry a 5% uncertainty, introducing an error of $\pm 0.35 \text{ Gt C yr}^{-1}$. Third, unlike the CO_2 – CH_4 covariation, the effect of the up-hole O_2 flux is sensitive to the assumed diffusivity–depth profile. Imposing a $\pm 5\%$ change in the modelled diffusivity changes the land carbon flux by $\pm 0.5 \text{ Gt C yr}^{-1}$. Although this perturbation of 5% is somewhat arbitrary, it is sufficient to substantially degrade the agreement between the predicted and observed CO_2 –depth profiles. Fourth, we have assumed that the entire upward flux of O_2 originates at the top of the lock-in zone. In reality, the flux originates wherever bubbles are closing. At the South Pole, bubbles are sealed anywhere between ~100 m depth (where the density reaches 0.8 g cm^{-3} ; ref. 29) and total close-off 20 m below. The effect of our assumption about the depth of O_2 injection is impossible to quantify at present.

Quadratically adding the errors due to fitting, stoichiometry and diffusivities gives a final result of $0.4 \pm 1.1 \text{ Gt C yr}^{-1}$ released by the land biosphere. Based on the CO_2 values for the samples from 113 to 60 m depth, our samples reflect atmospheric O_2 concentrations from 1974 to 1985 (scaled for relative diffusivities of O_2 and CO_2). However, the deepest samples are most sensitive to the up-hole flux. Thus, we believe our fit reflects the carbon balance of

the land biosphere most robustly during the years 1977–85. Sensitivity studies confirm that our estimate is largely immune to changes in carbon storage by the land biosphere before 1974 and after 1988. The present results improve on, and supersede, the earlier work of Bender *et al.*²² based on Vostok firn air.

Our estimate of terrestrial carbon release is in good agreement with completely independent results of previous studies based on tracer-calibrated box-diffusion models and three-dimensional ocean general circulation models^{4,5}. Our observation of a balanced biosphere also implies a change in the carbon cycle between our measurement period (1977–85) and the period after 1990, when direct atmospheric measurements of $\delta^{13}C$ and O_2/N_2 indicate a substantial terrestrial sink of CO_2 (refs 6–8).

Atmospheric N_2O

The pre-industrial mixing ratio of atmospheric N_2O is not well defined. Estimates of pre-industrial levels of N_2O are based on ice-core measurements (see, for example, ref. 30) which do not extend forward beyond 1966, and exhibit substantial scatter. Estimates of the growth rate of atmospheric N_2O since the mid-1970s range from 0.52 to $1.0 \text{ nmol mol}^{-1}$ (see, for example, refs 31–34). Because the growth rate is slow and the earliest real-time measurements date only from the mid-1970s, these must still be considered short-term estimates. The firn data presented here provide a precise record of atmospheric N_2O concentration throughout this century and agree with contemporary sampling programs where the records overlap. Compared to the ice-core records, the higher precision and accuracy of the firn data should aid considerably in assessing the causes of the rising abundance of N_2O (ref. 35).

At the bottom of the profile, the dry mole fraction of N_2O is $278.8 \text{ nmol mol}^{-1}$, dating from 1903 (Figs 1, 3). This value, and others from deep in the firn, are 5–10 nmol mol^{-1} lower than corresponding ice-core data from a variety of cores^{33,36–40}, ~10 nmol mol^{-1} higher than ice-core measurements from South Greenland⁴¹, and in good agreement with the recent Antarctic core of Machida *et al.*⁴². Based on the firn air data, the atmospheric mole fraction of N_2O increased slowly during the first half of the century, at $0.06 \pm 0.01\% \text{ yr}^{-1}$ (95% confidence limits), reaching $288 \text{ nmol mol}^{-1}$ by 1958. Thereafter, N_2O increased much more rapidly, at $0.22 \pm 0.02\% \text{ yr}^{-1}$, to contemporary Southern Hemisphere values of $311 \text{ nmol mol}^{-1}$. The firn data

BOX 1 Equations and assumptions in the firn air composition model

Assumed atmospheric concentration histories.

CO_2 history: direct atmospheric sampling⁴⁵ and spline fit to Law Dome ice-core record²⁵.

CH_4 history: direct atmospheric sampling⁴⁶ and spline fit to Law Dome ice-core record. Law Dome record is a revision of ref. 24 reflecting minor changes in chronology and improved analytical precision (D. Etheridge, personal communication).

Model structure.

Transport equation: $\Delta C_{ij}/\Delta t \approx (1/S_i \Delta z^2)[D_{i-1/2} S_{i-1/2}(C_{i-1,j} - C_{i,j}) - D_{i+1/2} S_{i+1/2}(C_{i,j} - C_{i+1,j})]$ where S is open porosity, D is diffusivity, C is concentration, i is box number and j is the time-step index.

Time step: 4.8 hours.

Vertical divisions: 115 boxes, each 1 m thick.

Length of run: 1 January 1887 to 29 January 1995.

Porosity–depth profile: $S = 1 - \rho_{\text{fm}}/\rho_{\text{ice}}$ where S is open porosity and ρ is density²⁹.

Diffusivity–depth profile: Start with $D = 1.16 \text{ m}^2 \text{ d}^{-1}$ ($23.7S - 2.84$) (ref. 29) and adjust until model reproduces CO_2 –depth profile.

Diffusivities: $D_{O_2} = 1.27D_{CO_2}$ (ref. 29); $D_{CH_4} = 1.35D_{CO_2}$ (ref. 47).

Upward O_2 flux.

Flux: flux (F) = accumulation $\times$ fractional gas volume at close-off $\times O_2/N_2$ anomaly.

Accumulation and fractional gas volume: $7 \text{ g cm}^{-2} \text{ yr}^{-1}$ (ref. 20) and $0.0925 \text{ cm}^3 \text{ g}^{-1}$ (ref. 23), respectively.

Concentration anomaly: $C_{z^*} \approx C_{115\text{mbs}} - F \sum_{z=115\text{mbs}}^z (\Delta z/D(z))$ where C_{z^*} is the concentration change (from ambient) at depth z^* metres below surface (m.b.s.), due to O_2 flux only.

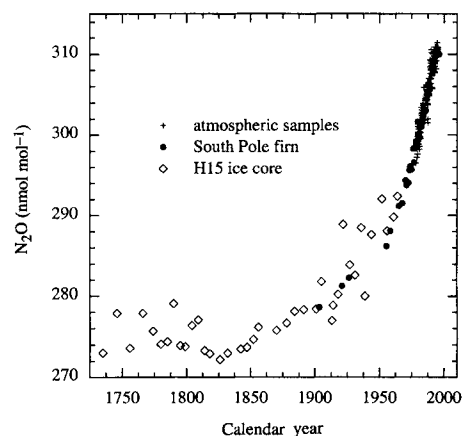


FIG. 3 History of atmospheric N_2O over the past 250 years, derived from Antarctic ice-core measurements⁴², real-time air measurements in the Southern Hemisphere³⁴ and analyses of South Pole firn air (this study). Firn-air ages are determined from correlation of CO_2 in the samples with the atmospheric CO_2 history of Etheridge *et al.*²⁵. Diffusivities of N_2O and CO_2 are assumed to be identical⁴⁷ since the two gases are isobaric and isoelectronic. Thus, CO_2 ages apply directly to N_2O . Ice-core data of Machida *et al.*⁴² have been lowered by 1 nmol mol^{-1} to conform to the NOAA/CMDL scale.

(12 points) and real-time measurements by NOAA/CMDL³⁴ show similar growth rates since 1978 (0.66 ± 0.11 and 0.75 ± 0.03 nmol mol⁻¹ yr⁻¹, respectively).

Covariation of N₂O with CO₂ is surprisingly linear throughout the profile (Fig. 2), with a slope of 0.50 ± 0.03 nmol N₂O per μ mol CO₂ (95% c.l., $r^2 = 0.98$). This result suggests that CO₂ and N₂O have been rising at a constant ratio ($dN_2O/dCO_2 = 2,000$). The

linearity may be coincidental, as N₂O and CO₂ do not share dominant sources and sinks. The atmospheric CO₂ budget has been outlined above, but most of the increase in atmospheric N₂O is currently unexplained. Direct emissions of N₂O from fossil-fuel burning are probably small. Other known sources include the ocean, fertilizers and natural soils. Atmospheric production *in situ* has also been suggested as a possibly significant source^{43,44}. □

Received 20 February; accepted 13 August 1996.

- Delmas, R. *Tellus* **45B**, 391–396 (1993).
- Schwander, J. et al. *J. Geophys. Res.* **98**, 2831–2838 (1993).
- Schwander, J. in *The Environmental Record in Glaciers and Ice Sheets* (eds Oeschger, H. & Langway, C.) 53–67 (Wiley, New York, 1989).
- Keeling, C. D. et al. in *Aspects of Climate Variability in the Pacific and the Western Americas* (ed. Peterson, D. H.) 165–236 (Geophys. Monogr. 55, Am. Geophysical Union, Washington, D.C., 1989).
- Sarmiento, J. L., Orr, J. C. & Siegenthaler, U. *J. Geophys. Res.* **97**, 3621–3645 (1992).
- Ciais, P., Tans, P., Trolier, M., White, J. W. C. & Frakey, R. J. *Science* **269**, 1098–1102 (1995).
- Bender, M., Ellis, T., Tans, P., Francey, R. & Lowe, D. *Glob. Biogeochem. Cycles* **10**, 9–21 (1996).
- Keeling, R. F., Piper, S. C. & Heimann, M. *Nature* **381**, 218–221 (1996).
- Colbeck, S. C. *J. Glaciol.* **35**, 209–213 (1989).
- Gibbs, J. W. *Collected Works Vol. 1, Thermodynamics* (Yale Univ. Press, New Haven, 1928).
- Craig, H., Horibe, Y. & Sowers, T. *Science* **242**, 1675–1678 (1988).
- Sowers, T., Bender, M. & Raynaud, D. *J. Geophys. Res.* **94**, 5137–5150 (1989).
- Herron, M. M. & Langway, C. C. *J. Glaciol.* **25**, 373–384 (1980).
- Bender, M., Sowers, T. & Lipenkov, V. J. *Geophys. Res.* **100**, 18651–18660 (1995).
- Grew, K. E. & Ibb, T. L. *Thermal Diffusion in Gases* (Cambridge Univ. Press, 1952).
- Severinghaus, J. P., Bender, M. L., Keeling, R. F. & Broecker, W. S. *Geochim. Cosmochim. Acta* (in the press).
- Severinghaus, J. P. & Sowers, T. (abstr) *Eos* **76**(46) November (1995).
- Patterson, W. *The Physics of Glaciers* (Pergamon, Oxford, 1969).
- Jones, P., Wigley, T. & Briffa, K. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. W.) 603–608 (No. ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge, 1994).
- Gow, A. J. *J. Glaciol.* **5**, 467–477 (1965).
- Kuvinen, K. C., Koci, B. R., Holdsworth, G. W. & Gow, A. J. *Antarct. J. US* **17**, 89–91 (1982).
- Bender, M. L., Sowers, T., Barnola, J. M. & Chapellaz, J. *Geophys. Res. Lett.* **21**, 189–192 (1994).
- Martinerie, P., Raynaud, D., Etheridge, D. M., Barnola, J.-M. & Mazaudier, D. *Earth Planet. Sci. Lett.* **112**, 1–13 (1992).
- Etheridge, D. M., Pearman, G. & Fraser, P. *Tellus* **44b**, 282–294 (1992).
- Etheridge, D. M. et al. *J. Geophys. Res.* **101**, 4115–4128 (1996).
- Machta, L. in *Proc. Int. Meeting on Stable Isotopes in Tree Ring Research* (ed. Jacoby, G. C.) 125–127 (1980).
- Keeling, R. F. thesis, Harvard Univ. (1988).
- Keeling, R. F. & Shertz, S. R. *Nature* **358**, 723–727 (1992).
- Schwander, J., Stauffer, B. & Sigg, A. *Ann. Glaciol.* **10**, 141–145 (1988).
- Etheridge, D., Pearman, G. & deSilva, F. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. W.) 391–394 (No. ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge, 1994).
- Weiss, R. J. *Geophys. Res.* **86**, 7185–7195 (1981).
- Prinn, R. et al. *J. Geophys. Res.* **95**, 18369–18385 (1990).
- Khalil, M. A. K. & Rasmussen, R. J. *Geophys. Res.* **97**, 14651–14660 (1992).
- Thompson, T. M. et al. in *Climate Monitoring and Diagnostics Laboratory No. 22 Summary Report 1993* (eds Peterson, J. & Rossen, R.) 72–91 (US Dept. of Commerce, Boulder, Colorado, 1994).
- Nevison, C., Esser, G. & Holland, E. *Clim. Change* **32**, 327–378 (1996).
- Khalil, M. *Encyclopedia of Environmental Biology* Vol. 2, 251–265 (Academic, New York, 1995).
- Dobb, J., Rasmussen, R., Mayewski, P. & Holdsworth, G. *Chemosphere* **27**, 2413–2423 (1993).
- Zardini, D., Raynaud, D., Scharffe, D. & Seiler, W. *J. Atmos. Chem.* **8**, 189–201 (1989).
- Khalil, M. & Rasmussen, R. *Ann. Glaciol.* **10**, 73–79 (1988).
- Pearman, G., Etheridge, D., de Silva, F. & Fraser, P. *Nature* **320**, 348–350 (1986).
- Leuenberger, M. & Siegenthaler, U. *Nature* **360**, 449–451 (1992).
- Machida, T., Nakazawa, T., Fujii, Y., Aoki, S. & Watanabe, O. *Geophys. Res. Lett.* **22**, 2921–2924 (1995).
- Prasad, S. J. *Geophys. Res.* **99**, 5285–5294 (1994).
- Schade, G. & Crutzen, P. J. *Atmos. Chem.* **22**, 319–346 (1995).
- Thoning, K. W., Tans, P. P. & Waterman, L. S. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. W.) 28–40 (No. ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge, 1994).
- Diugokenky, E. J., Lang, P. M., Masarie, K. A. & Steele, L. P. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. W.) 262–266 (No. ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge, 1994).
- Landolt-Börnstein, *Eigenschaften der Materie in ihren Aggregatzuständen* Vol. 6, Ch 5a (Springer, Berlin, 1969).
- Conway, T. et al. *Tellus* **40B**, 81–115 (1988).
- Butler, J., Elkins, J., Thompson, T. & Egan, K. J. *Geophys. Res.* **94**, 14865–14877 (1989).
- Bender, M. L., Tans, P. P., Ellis, J. T., Orchard, J. & Halbasset, K. *Geochim. Cosmochim. Acta* **58**, 4751–4758 (1994).
- Mariand, G., Andres, R. J. & Boden, T. A. in *Trends '93: A Compendium of Data on Global Change* (eds Boden, T. A., Kaiser, D. P., Sepanski, R. J. & Stoss, F. W.) 504–584 (No. ORNL/CDIAC-65, Carbon Dioxide Information Analysis Center, Oak Ridge, 1994).

ACKNOWLEDGEMENTS. We thank D. Giles and J. Kyne for carefully drilling the holes from which these samples were drawn, and D. Etheridge for discussions. This work was supported by the Office of Polar Programs (US NSF), the National Institute of Global Environmental Change Northeast Regional Center (US Department of Energy), the US Environmental Protection Agency, and the Atmospheric Chemistry Project of NOAA's Climate and Global Change Research Program.

CORRESPONDENCE should be addressed to M.B. (e-mail: battle@gsosun1.gso.uri.edu).

KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Galaxy formation at high redshifts

N. Metcalfe, T. Shanks, A. Campos*, R. Fong & J. P. Gardner

Physics Department, University of Durham, South Road, Durham DH1 3LE, UK

* Present address: Instituto de Astrofísica de Andalucía, CSIC, Spain.

SENSITIVE optical surveys have revealed¹ a large population of 'faint blue galaxies', which are believed to be young galaxies observed close to their time of formation². But there has been considerable uncertainty regarding the epochs at which these galaxies are observed, owing to the difficulties inherent in determining spectroscopic redshifts for very faint objects. Here, by modelling the numbers and colours of galaxies at the faintest detection limits, we show that the faint blue galaxies are likely to lie at high redshift ($z \approx 2$). This conclusion holds regardless of whether the Universe is assumed to be open or at the critical density (flat). In an open universe, the data are consistent with galaxy models in which star formation rates decay exponentially with decreasing redshift, whereas the assumption of a flat universe requires the addition of a population of galaxies which are seen only at high redshift.

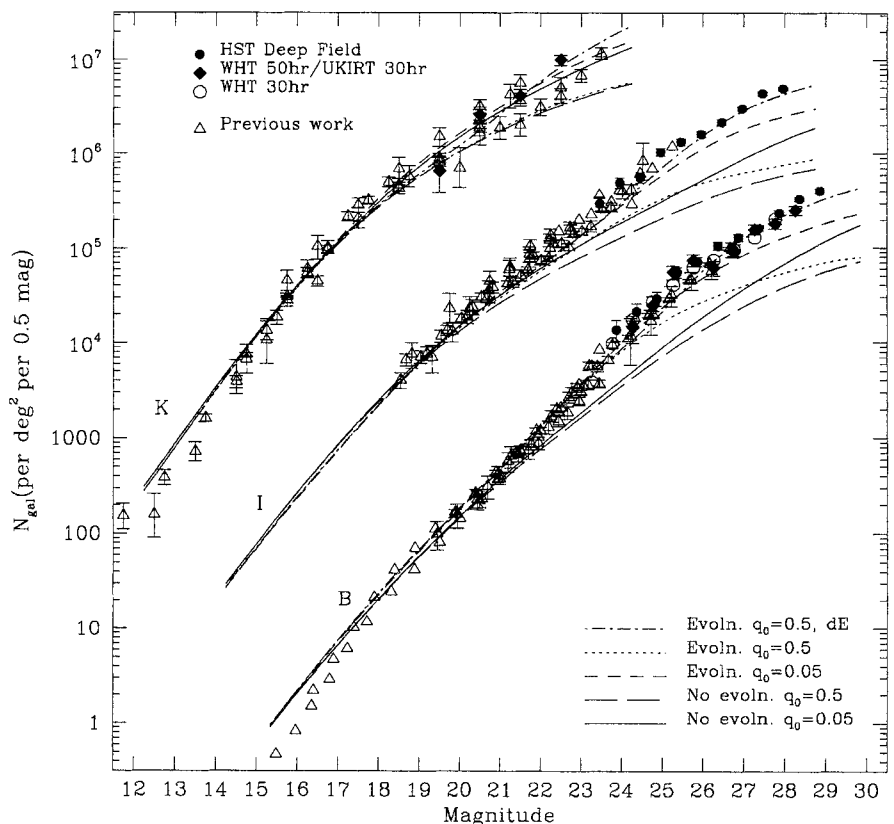
We have measured galaxy counts and colours in the Deep Field³ recently observed by the Hubble Space Telescope (HST) to $U = 27$ mag, $B = 29$ mag, $R = 28.5$ mag and $I = 28$ mag. The HST data covers 5.3 arcmin^2 and the exposure times were ~ 25 h

in each of the BRI bands and ~ 47 h in the U band. Similar techniques, adjusted for the smaller image size, were used to analyse the HST data as we have previously used to analyse deep ground-based images^{4,5}. Taking these deepest galaxy count data together with the results of recent galaxy redshift surveys at the spectroscopic limit of the Keck Telescope^{6,7}, reveals the clearest picture so far of the Universe at faint magnitudes.

The derived HST B galaxy number counts are shown in Fig. 1, together with those from other work, including the deepest ground-based counts from the William Herschel Telescope (WHT) to $B = 28.2$ mag. The HST and WHT results are seen to be in good agreement to this limit, with the HST counts extending ~ 1 mag fainter. Figure 1 also shows the HST counts in the I band. In this band the HST data extends about ten times deeper than previous data, because of the higher HST resolution and the fainter background sky. Although HST has as yet no K-band imaging capability, we also summarise in Fig. 1 the deepest ground-based K counts, including new data from the UK Infrared Telescope (UKIRT). The K counts now appear to be reasonably well defined in the range $13 \text{ mag} < K < 24 \text{ mag}$.

We first compare the count data with non-evolving models. In the modelling we assume throughout a Hubble constant of $H_0 = 50 \text{ km s}^{-1} \text{ Mpc}^{-1}$, but our main conclusions below are insensitive to this assumption. As can be seen from Fig. 1, the non-evolving models increasingly underestimate the optical counts at faint magnitudes. However, as noted previously, the magnitude at which the 'faint blue galaxy' excess becomes apparent is dependent on the normalization of the models at bright magnitudes. With the high normalization adopted here, non-evolving models give a reasonable representation of the bright counts and redshift distributions in the range $18 \text{ mag} < B < 22.5 \text{ mag}$ (refs 1, 4, 5, 8, 9)

FIG. 1 The B (Johnson) and I (Kron-Cousins) galaxy counts from the HST Deep Field (F450W, F814W) data compared to counts from the WHT, UKIRT and elsewhere (refs 1, 4, 5, 10, 11, 22–29). Also shown are deep K-band galaxy counts from ground-based data. The I-band counts have been multiplied by a factor of 10, and the K-band counts have been multiplied by a factor of 100, for clarity. (The HST U and R counts are available: see Supplementary Information.) The model luminosity function parameters and other details of the modelling procedure are given elsewhere^{4,5}. The galaxy luminosity evolution with redshift is computed from Bruzual and Charlot¹² isochrone synthesis models, using the appropriate passbands. We assume galaxy ages of 16 and 12.7 Gyr in the cases $q_0 = 0.05$, 0.5. We adopt a dwarf-dominated IMF ($x = 3$) with an exponentially increasing star-formation rate of timescale, $\tau = 2.5$ Gyr, for E/S0/Sab galaxies, and a Salpeter IMF ($x = 1.35$) with an exponentially increasing star-formation rate of time scale, $\tau = 9$ Gyr, for Sbc/Scd/Sdm galaxies. The latter are also assumed to have internal dust absorption at $z = 0$ of $A_B = 0.3$ mag, $A_I = 0.11$ mag, $A_K = 0.03$ mag. We take an absorption law inversely proportional to the wavelength λ , $A_\lambda \propto 1/\lambda$, to approximate the effect of redshift on the internal dust absorption. The models also include the effect of Lyman- α forest/break absorption²¹. Spiral evolution dominates these models in the B and I bands. The $q_0 = 0.05$ evolving model gives a good fit to $B \approx 27$ mag, $I \approx 26$ mag, whereas the $q_0 = 0.5$ model only fits to $B \approx 25$ mag, $I \approx 23.5$ mag. The fit of the $q_0 = 0.5$ model is improved when an extra high redshift galaxy population (dE) with constant star formation rate (SFR) at $z > 1$ and rapidly fading at $z < 1$ is invoked. The Schechter luminosity function parameters of



the dE population at $z = 0$ are $M_B^* = -16.0$ mag, $x = -1.2$ and $\phi^* = 0.019 \text{ mag}^{-1} \text{ Mpc}^{-3}$. The K galaxy counts, in contrast to the B and I counts, are well fitted by non-evolving models.

and also to individual counts of spiral and early-type galaxies at similar magnitudes^{10,11}. As a consequence, the B galaxy count then only shows evidence for strong evolution in the range $B > 23$ mag. Also the high normalization allows non-evolving models with deceleration parameter, $0.05 < q_0 < 0.5$ to fit the K counts from $K = 15$ mag to the faintest count limit at $K = 24$ mag. This is as expected, since evolved, young star populations are expected to affect galaxy light less in the infrared than in the blue.

To fit the optical counts, we then consider simple evolutionary models from Bruzual and Charlot¹² where galaxy star-formation rates rise exponentially with look-back time. In such models, the galaxy light at early times is generally dominated by luminous, blue stars but at later times, when these stars fade and the star-formation rate slows, the galaxy light dims and reddens. These models are known to fit the B counts in the range $18 \text{ mag} < B < 25 \text{ mag}$ ^{1,4,5,8,9}. But previous faint galaxy redshift surveys at $B < 24$ mag presented a problem for such models, as they predict a high-redshift tail of evolved, luminous galaxies which was unobserved in these surveys¹³. However, these surveys were frequently only $\sim 60\%$ complete, and the galaxies with unidentified redshifts were usually blue. Recently Cowie *et al.*^{6,7} have used the Keck 10-m telescope to make a new $B < 24$ mag galaxy redshift survey with $> 80\%$ completeness and have detected such a high-redshift galaxy component, supporting the basic viability of these models (see Fig. 2). An extended high-redshift tail is also consistent with the low galaxy clustering amplitude observed⁹ at $B > 23$ mag.

A further new development is that the latest version of the preferred spiral galaxy evolution model¹² now permits more brightening at high redshift, while still reproducing the blue colours of spirals at the present day¹⁴. Combined with the steep luminosity function of spirals⁸ and assuming a small amount of internal dust¹⁴⁻¹⁶, this enables us to obtain a good fit to the high-redshift tail of the Keck number-redshift distribution, $n(z)$, at $22.5 \text{ mag} < B < 24 \text{ mag}$ (see Fig. 2) and to other redshift survey results¹⁷. In the low- q_0 case, this spiral-dominated model then also produces a reasonable fit to the optical counts to $B \approx 27$ mag and $I \approx 26$ mag (see Fig. 1). Although the model

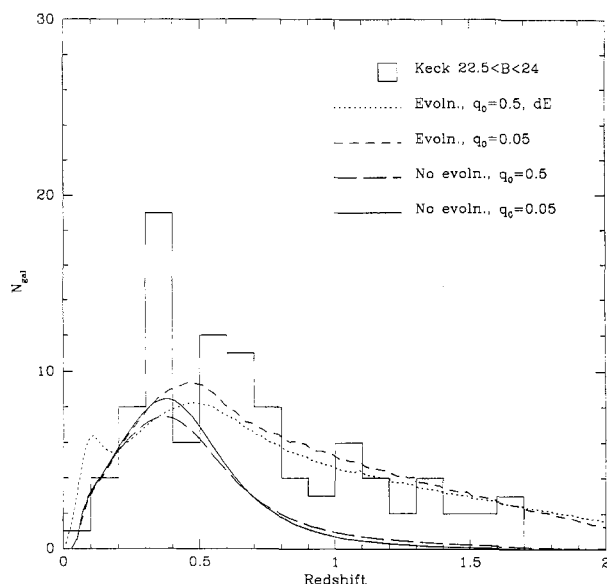
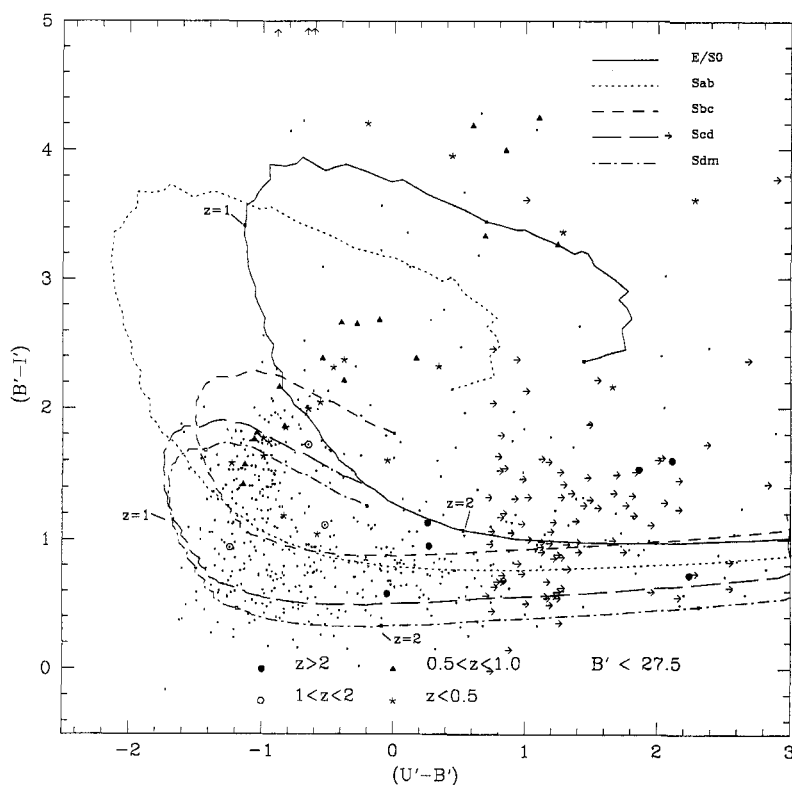


FIG. 2 The galaxy number-redshift distribution, $n(z)$, for $22.5 \text{ mag} < B < 24 \text{ mag}$ implied by new redshift data acquired on the Keck Telescope (refs. 6, 7). The observed $n(z)$ is clearly more extended than the non-evolving models with either $q_0 = 0.05$ or $q_0 = 0.5$. The extended redshift distribution is well fitted by our evolutionary models whose parameters are described in Fig. 1 legend.

underestimates the optical counts at fainter magnitudes, this discrepancy is probably still within the combined data and model uncertainties. In the $q_0 = 0.5$ case, the spiral luminosity evolution model only fits the optical data to $B \approx 25$ mag and $I \approx 23.5$ mag and then more seriously underestimates the counts at fainter magnitudes. Thus, the HST data confirms the previous

FIG. 3 Dots represent the $U' - B'$: $B' - I'$ colours of galaxies with $B' < 27.5$ mag in the Hubble Deep Field. Primed letters for magnitudes indicate that here we are using the natural HST magnitude system, with the zero point set at an AOV star. The arrows represent detection upper limits, mainly galaxies which are undetected in U' . The $U' - B'$ colours move sharply redwards at $B' - I' \approx 0.8$ due to the Lyman- α forest/Lyman break passing through the U' band. The predicted tracks are the $q_0 = 0.05$ evolutionary models for each morphological type as detailed in Fig. 1 legend, modulated in the case of Sbc/Scd/Sdm types by our assumed internal dust absorption of $A_U = 0.45 \text{ mag}$, $A_B = 0.3 \text{ mag}$, $A_I = 0.11 \text{ mag}$ and in the case of all galaxies by the Lyman- α forest absorption. The models used in the $q_0 = 0.5$ case (not shown) show a very similar behaviour, even for the rapidly fading dE type. The $z = 1$ and $z = 2$ labelled positions on the tracks indicate the colours of model E/SO and Sdm galaxies at these redshifts. The remaining symbols show the colours of 45 brighter galaxies with Keck spectroscopic redshifts, and these agree well with the predicted colours for these galaxies. It can also be seen that $U' - B' < 0$ is predicted to correspond to galaxies with $z < 2$, and $U' - B' > 0$ to galaxies with $z > 2$.



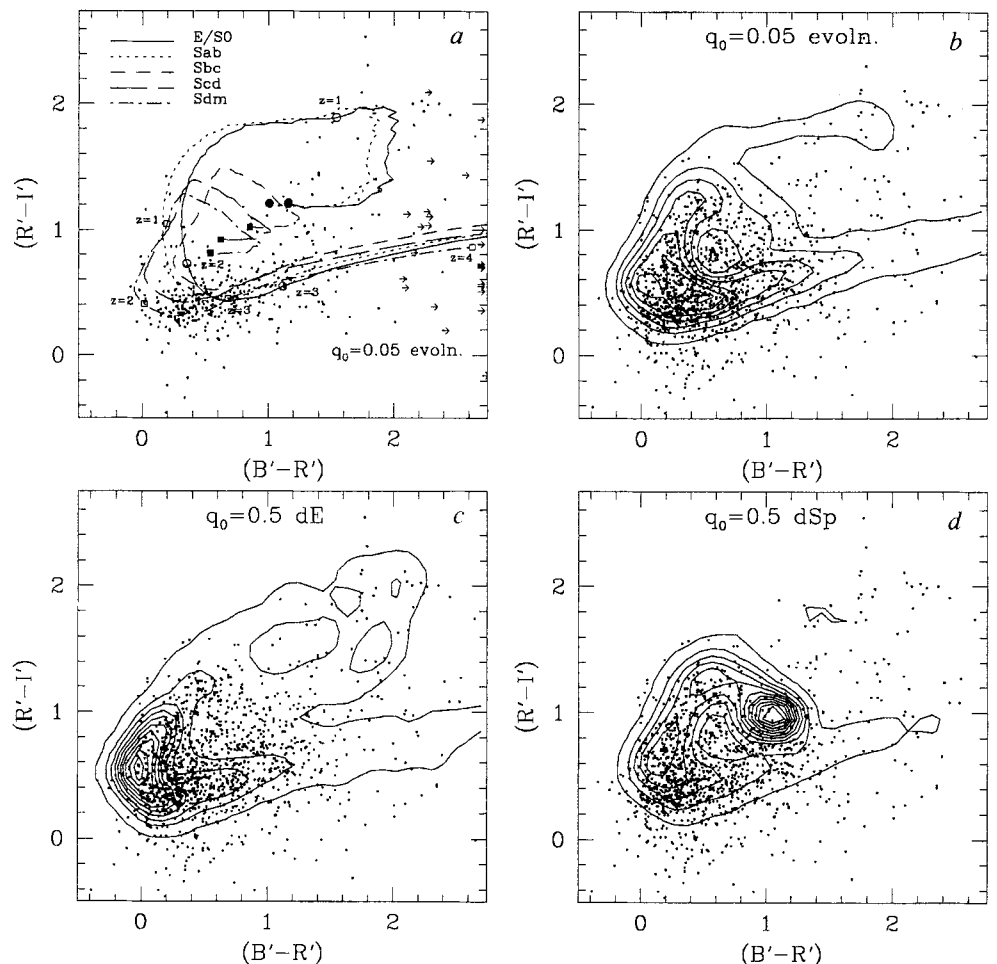
ground-based result^{1,4,5,18} that if $q_0 = 0.5$, then there is not enough spatial volume at high redshifts to allow simple luminosity evolution models to fit the high galaxy counts at $B > 25$ mag. This confirmation is important because confusion corrections are much smaller in the high-resolution HST galaxy counts than in previous ground-based data.

We have already noted that the faint K counts seem less affected by evolution and this is supported by the new Keck redshift data^{6,7}, where non-evolving models again give a good fit to the $n(z)$ relation for $18 \text{ mag} < K < 19 \text{ mag}$ (ref. 7). As non-evolving models imply that, at $K < 20 \text{ mag}$, the majority of galaxies are early types, this suggests that early-type galaxies may be little affected by either dynamical (merging) or luminosity evolution. Indeed, although the effect of passive evolution of early-type galaxies with a Salpeter initial mass function (IMF) slope ($\alpha = 1.35$) is small at K ($< 0.7 \text{ mag}$ at $z = 1$), it is still big enough to make the predicted Keck $n(z)$ appear too extended at $18 \text{ mag} < K < 19 \text{ mag}$ (refs 6, 7). If the result is not due to incompleteness, then the amount of K evolution can be reduced to acceptable limits by assuming a more dwarf-dominated IMF ($\alpha = 3$) for early-type galaxies. While awaiting more complete K surveys, we have adopted this dwarf-dominated IMF for early types throughout this paper.

To improve the fit of the $q_0 = 0.5$ model to the faint optical counts, we consider a model with an extra population of high-redshift galaxies which have a constant star-formation rate from the formation epoch till $z = 1$ ($\sim 4 \text{ Gyr}$ after formation with our assumed H_0). The Bruzual and Charlot¹² model shows that at $z \lesssim 1$ the galaxy then rapidly fades by 5 magnitudes in B to form a red (dE) galaxy by the present day. This model is in the spirit of previously proposed 'disappearing dwarf' count models¹⁹ and it gives a good fit to the faint B, I and K counts and the Keck $n(z)$ data (see Figs 1 and 2).

We next test these models against the faint galaxy colour distributions in the Hubble Deep Field. The presence of broad features in galaxy spectra allows tests to be made of the predicted galaxy redshifts. One approach is to use the HST broad-band photometry as a rough galaxy spectrum and then derive redshifts for individual galaxies using local galaxies as templates²⁰. Here we follow the different approach of simply comparing our evolutionary model predictions to the observations of faint galaxy colours. We see in Fig. 3 that our predicted $U' - B' : B' - I'$ model tracks, that is, the loci traced by model galaxy colours as they change with redshift, compare well with the observed colours for $B' < 27.5 \text{ mag}$ galaxies. (Primed letters here denote the natural HST magnitude system). In particular, the redshifting of the

FIG. 4a, The $q_0 = 0.05$ evolutionary models' $B' - R' : R' - I'$ tracks with redshift, as a function of galaxy morphological type. Primed letters for magnitudes indicate that here we are using the natural HST magnitude system, zero-pointed to an A0V star. The tracks are modulated by our assumed internal dust and Lyman- α forest/break absorption. A filled symbol marks each galaxy type's colour at zero redshift. The open symbols mark the galaxy type's colour at unit intervals in redshift for E/S0 and Sdm galaxies. Galaxies with $B' - R' = 0.3 \text{ mag}$ and $R' - I' = 0.6 \text{ mag}$ have $z \approx 2$. The same is true for the $q_0 = 0.5$ models, even for the rapidly fading dE type (not shown). The dots indicate the colours of galaxies with $R' < 27.5 \text{ mag}$ and $U' - B' > 0$ which are predicted to have $z > 2$. Their position in the independent $B' - R' : R' - I'$ plane is consistent with this prediction. b, Dots represent the $B' - R' : R' - I'$ colours of $R' < 28 \text{ mag}$ galaxies in the Hubble Deep Field. These show the same horse-shoe-shaped track as expected from the models in a. The contours represent the relative numbers of galaxies predicted for the $q_0 = 0.05$ evolutionary model, which appear to be in good agreement with the data, peaking at the colours corresponding to $z \approx 2$. c, As b, but with the contours here representing the relative numbers of galaxies predicted for the $q_0 = 0.5$ 'disappearing dwarf' (dE) model, which again appear to be in reasonable agreement with the data. d, As b but with the contours here representing the relative numbers of galaxies predicted for the $q_0 = 0.5$ low-redshift dwarf (dSp) model. The data show that this model is inappropriate, as it clearly predicts too many low-redshift, $z \approx 0.5$, spiral galaxies.



Lyman- α forest/break²¹ absorption features through the U' band causes the model $U' - B'$ colours to move sharply redwards at $B' - I' \approx 0.7$ mag and the same effect is clearly seen in the data. The colours of 45 brighter, $B \approx 24$ mag galaxies with Keck redshifts are also shown in Fig. 3, and are also found to agree well with their predicted colours. (The equivalent $B' - R'$ versus $R' - I'$ graph is available: see Supplementary Information).

The above predictions show that, for the majority of faint galaxies, $U' - B' < 0$ is predicted to correspond to $z < 2$ galaxies and $U' - B' > 0$ corresponds to $z > 2$ galaxies. We find that the proportion of galaxies with $U' - B' > 0$ (including those undetected in U') rises to $47 \pm 7\%$ of the total at $27 \text{ mag} < B < 28 \text{ mag}$, indicating that the redshift distribution may peak at $z \approx 2$. This fraction is matched very well by both the $q_0 = 0.05$ model which predicts 47% with $U' - B' > 0$ at the same limit and the $q_0 = 0.5$, disappearing dwarf (dE) model which predicts 43%. We have also considered another $q_0 = 0.5$ model which assumes an extra population of low-redshift dwarf spirals (dSp) which evolve more slowly according to our standard exponential model for spiral luminosity evolution¹². Although this model also gives an improved fit to the counts, it predicts too few high-redshift ($U' - B' > 0$) galaxies (28%) for compatibility with the faint HST data.

These conclusions are confirmed by consideration of the $B' - R'$ versus $R' - I'$ colour-colour plot in Fig. 4. Figure 4a shows the predicted tracks of the galaxy types with redshift, as in Fig. 3. Also plotted are the galaxies with $U' - B' > 0$ and $R' < 27.5$ mag; these are expected to have $z > 2$ by the above arguments and it can be seen that their position on the $B' - R' : R' - I'$ tracks is entirely consistent with their lying in this redshift range. We regard this as crucial confirmation that our models are indicating consistent redshifts for galaxies in $U' - B'$ and $B' - R' : R' - I'$ independently. Figure 4b, c, d then show the HST data (dots) at $R' < 28$ mag compared to the predicted galaxy number contours for the open and closed models, based on the tracks shown in Fig. 4a. Both the $q_0 = 0.05$ and the $q_0 = 0.05$ dE model contours give a reasonable fit to the data which seem to peak at $B' - R' \approx 0.3$, $R' - I' \approx 0.3$, corresponding to $z \approx 2$ for all galaxy types. However, the $q_0 = 0.05$ dSp model contours peak away from this point at $B' - R' \approx 1$, $R' - I' \approx 1$ which corresponds to $z \approx 0.5$ for the dwarf spiral galaxies and we conclude that the galaxy redshift distribution in this model is skewed to too low redshifts to be compatible with the colour data. This does not mean that the above 'disappearing dwarf' model is unique in allowing a fit to be obtained with $q_0 = 0.5$; other possibilities such as merging models may also exist. However, it does suggest that, in any model, the star-forming phase has to be at $z \approx 2$ for consistency with the faint galaxy colours in the Hubble Deep Field. $\square$

23. Driver, S. P., Philipps, S., Davies, J. I., Morgan, I. & Disney, M. J. *Mon. Not. R. Astron. Soc.* **268**, 393–404 (1994).
24. Djorgovski, S. et al. *Astrophys. J.* **438**, L13–L16 (1995).
25. Gardner, J. P., Cowie, L. L. & Wainscoat, R. J. *Astrophys. J.* **415**, L9–L12 (1993).
26. Soffer, B. T. et al. *Astrophys. J.* **420**, L1–L4 (1994).
27. McLeod, B. A., Bernstein, G. M., Rieke, M. J., Tollestrup, E. V. & Fazio, G. G. *Astrophys. J. Suppl. Ser.* **96**, 117–121 (1995).
28. Glazebrook, K., Peacock, J. A., Collins, C. A. & Miller, L. *Mon. Not. R. Astron. Soc.* **266**, 65–91 (1995).
29. Gardner, J. P., Sharples, R. M., Carrasco, B. E. & Frenk, C. S. *Mon. Not. R. Astron. Soc.* (submitted).

SUPPLEMENTARY INFORMATION. Available on Nature's World-Wide Web site <http://www.nature.com> or as paper copies from Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank L. L. Cowie for allowing us to use results from the Keck 10-m redshift surveys in advance of publication, A. G. Bruzual for producing dwarf-dominated evolutionary models specially for this Letter, and the referees for helpful comments. We acknowledge the use of the Hubble Space Telescope Deep Field data. A.C. was supported by an EC Fellowship, and N.M. and J.P.G. by the UK PPARC.

CORRESPONDENCE should be addressed to T.S. (e-mail: Tom.Shanks@durham.ac.uk).

Entropic control of particle motion using passive surface microstructures

A. D. Dinsmore*, A. G. Yodh* & D. J. Pine†

* Department of Physics and Astronomy, University of Pennsylvania, 209 South 33 Street, Philadelphia, Pennsylvania 19104, USA

† Departments of Chemical Engineering and Materials, University of California, Santa Barbara, California 93106, USA

In a colloidal suspension containing particles of two different sizes, there is an attractive force between the larger particles. This attraction is due to the extra volume that becomes available to the smaller particles when the larger particles approach one another, thus increasing the entropy of the system. Entropic 'excluded-volume' effects of this type have been studied previously in colloids and emulsions, in the context of phase-separation phenomena in the bulk^{1–15} and at flat surfaces^{2,16}. Here we show how similar effects can be used to position the larger particles of a binary mixture on a substrate, or to move them in a predetermined way. Our experiments demonstrate the entropically driven repulsion of a colloidal sphere (in a suspension of smaller spheres) from the edge of a step; the magnitude of the entropic barrier felt by the sphere is approximately twice its mean thermal energy. These results indicate that passive structures etched into the walls of a container create localized entropic force fields which can trap, repel or induce the controlled drift of particles. Manipulation techniques based on these effects should be useful for making the highly ordered particle arrays required for structures with photonic band gaps^{17,18}, microelectronic mask materials¹⁹, and materials for clinical assays²⁰.

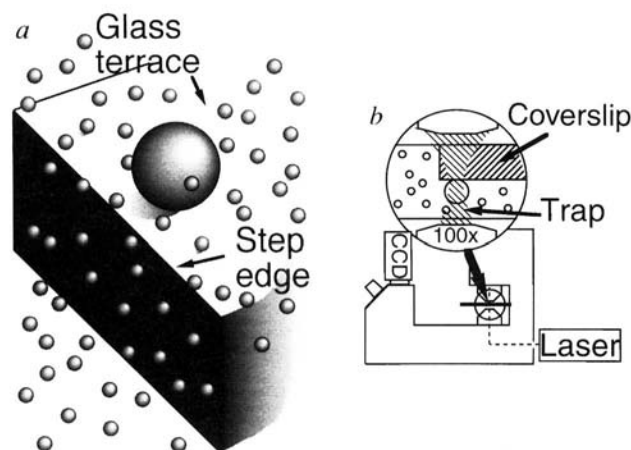
To demonstrate the entropic force-field idea in its barest form, we investigated the motions of hard spheres near the edge of a terrace (Fig. 1). We used an aqueous suspension of spherical polystyrene particles (Seradyn, Inc., Indianapolis) with diameters 0.460 μm and 0.083 μm and volume fractions 10^{-5} and 0.30, respectively. NaCl (0.01 M) was added to screen the electrostatic interactions over a distance of ~ 5 nm to obtain nearly ideal hard-core interactions¹⁶. As discussed below, the small spheres induce a 'depletion' attraction between the large spheres and the flat surface. Thus, a large sphere placed on the terrace diffused on its surface for several seconds (before escaping to the bulk). By following the two-dimensional trajectories of these spheres, we measured the forces acting on them near the step edge.

We quantify the effect of the step edge by measuring directly the Helmholtz free energy, F , of the system as a function of the position of the large sphere (Fig. 2a); we follow the analytical procedure given in ref. 21. Using video microscopy, we deter-

Received 12 February; accepted 13 August 1996.

1. Koo, D. C. & Kron, R. G. *Ann. Rev. Astron. Astrophys.* **30**, 613–652 (1992).
2. Cowie, L. L. In *The Post-Recombination Universe* (eds Kaiser, N. & Lasenby, A.) 1–18 (Kluwer, Dordrecht, 1990).
3. Williams, R. E. et al. *Astron. J.* (in the press).
4. Metcalfe, N., Shanks, T., Fong, R. & Jones, L. R. *Mon. Not. R. Astron. Soc.* **249**, 481–497 (1991).
5. Metcalfe, N., Shanks, T., Fong, R. & Roche, N. *Mon. Not. R. Astron. Soc.* **273**, 257–276 (1995).
6. Cowie, L. L., Hu, E. M. & Songaila, A. *Nature* **377**, 603–605 (1995).
7. Cowie, L. L., Songaila, A., Hu, E. M. & Cohen, J. G. *Astron. J.* (in the press).
8. Shanks, T. In *The Galactic and Extragalactic Background Radiations* (eds Bowyer, S. & Leinert, C.) 269–281 (Kluwer, Dordrecht, 1990).
9. Roche, N., Shanks, T., Metcalfe, N. & Fong, R. *Mon. Not. R. Astron. Soc.* **263**, 360–368 (1993).
10. Glazebrook, K., Ellis, R. S., Santiago, B. & Griffiths, R. *Mon. Not. R. Astron. Soc.* **275**, L19–L22 (1995).
11. Driver, S. P. et al. *Astrophys. J.* **449**, L23–L28 (1995).
12. Bruzual, A. G. & Charlot, S. *Astrophys. J.* **405**, 538–553 (1993).
13. Glazebrook, K. et al. *Mon. Not. R. Astron. Soc.* **273**, 157–168 (1995).
14. Campos, A. & Shanks, T. *Mon. Not. R. Astron. Soc.* (submitted).
15. Wang, B. *Astrophys. J.* **383**, L37–L40 (1991).
16. Gronwall, C. & Koo, D. C. *Astrophys. J.* **440**, L1–L4 (1995).
17. Lily, S. J., Tresse, L., Hammer, F., Crampton, D. & Le Fevre, O. *Astrophys. J.* **455**, 108–124 (1995).
18. Yoshii, Y. & Takahara, F. *Astrophys. J.* **326**, 1–18 (1988).
19. Babul, A. & Rees, M. J. *Mon. Not. R. Astron. Soc.* **255**, 346–350 (1992).
20. Lanzetta, K. M., Yahil, A. & Fernandez-Soto, A. *Nature* **381**, 759–763 (1996).
21. Madou, P. *Astrophys. J.* **441**, 18–27 (1995).
22. Smail, I., Hogg, D. W., Yan, L. & Cohen, J. G. *Mon. Not. R. Astron. Soc.* (in the press).

FIG. 1 *a* Drawing of a single large sphere near the edge of a terrace, representing the geometry of our experiment. The surrounding small spheres cause the large sphere to move in two dimensions on the surface and to be repelled from the step edge. The effect of gravity was negligible owing to the small particle sizes. *b*, Schematic diagram of the experimental apparatus. The (transparent) sample cell, $\sim 100\text{ }\mu\text{m}$ thick, was placed in the microscope for viewing with a $100\times$ (numerical aperture = 1.3) oil-immersion objective with differential interference contrast (DIC). A polished-glass cover slip provided the smooth terrace and step edge inside the cell. Because ϕ_L (see text) was small, the large spheres were $10\text{--}20\text{ }\mu\text{m}$ apart and did not interact with one another. A laser trap ($1,053\text{ nm}$ wavelength radiation, 10 mW into the objective)^{28,29} was used to grab a single large sphere and place it on the terrace $0.56 \pm 0.06\text{ }\mu\text{m}$ apart from the step edge. The laser trap was then turned off and the motions of the diffusing large sphere were followed using a CCD (charge-coupled-device) camera. We repeated the process to obtain $\sim 1,400$ events. Images 0.1 s apart were processed and the centre-of-mass sphere positions were measured with a precision of $0.04\text{ }\mu\text{m}$. The total number of large spheres in view was $1,388$ after 0.1 s but decreased by 33% over 4 s owing to escape of the spheres vertically from the flat surface into the bulk. In obtaining these data, we used several different $0.46\text{-}\mu\text{m}$ spheres and different sample cells, and always obtained similar results.



mined the probabilities, P_{ij} , that the large sphere jumped from the j th to the i th spatial bin during a time interval of length 0.1 s . We then computed the eigenvector of the matrix P with unit eigenvalue. The i th component of this eigenvector is proportional to $e^{-F_i/k_B T}$, where F_i is the Helmholtz free energy evaluated at the i th bin. The force on the large sphere points toward decreasing F with magnitude given by the gradient of F . To show the effect of the step edge on particle motions, we also present the probability distribution of the large-sphere positions at two times: 0.1 and 4.1 s after release (Fig. 2*b, c*).

Our results clearly exhibit free diffusion along the direction parallel to the step edge. The free energy was constant, indicating no net force (Fig. 2*a*, inset). Furthermore, as shown in Fig. 2*b*, the mean position, $\bar{y}$, remained zero and the probability distribution remained symmetric about the starting position (zero). The mean-square position, $\bar{y}^2$, was proportional to time, as is characteristic of diffusion.

Perpendicular to the edge, however, there was a pronounced free-energy barrier approximately $2k_B T$ in magnitude, located at the step edge. As a large sphere moved towards the edge, the free

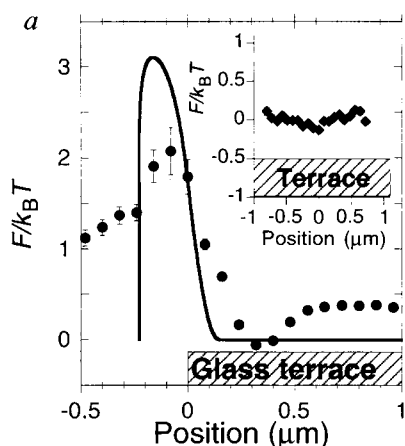
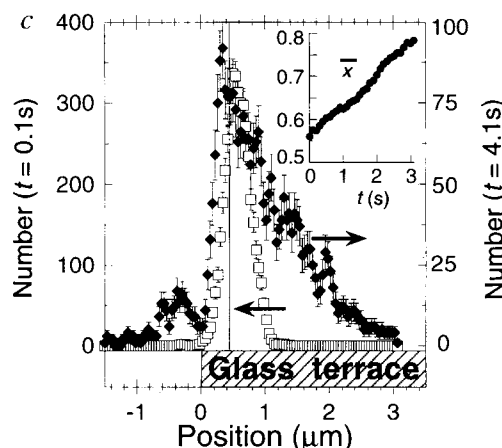
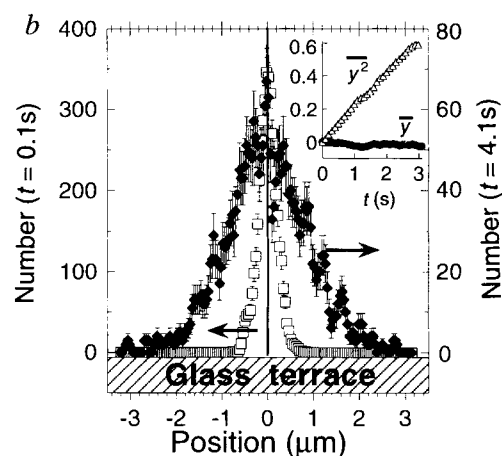


FIG. 2 *a*, Free energy of the system in units of the thermal energy, $F/(k_B T)$, as a function of the distance of a large sphere from a step edge located at zero. The filled circles represent measurements. The solid curve was calculated using the excluded-volume theory described in the text and treating the small spheres as an ideal gas with volume fraction 0.3 . Although the ideal-gas approximation underestimates the pressure of a hard-sphere fluid, we use it because other quantitative effects (for example, non-uniform distribution of small spheres) are also ignored. Inset, plot of $F/(k_B T)$ versus large-sphere position parallel to the edge. There is no barrier. *b*, Measured probability distributions of a single large sphere in the direction parallel to the step edge at two different times after release from position zero. Empty squares (left axis), 0.1 s ; filled diamonds (right axis), 4.1 s . The height of each data point gives the number of events in which the sphere was observed inside a $0.12\text{ }\mu\text{m}$ -wide bin centred at that point. Inset: upper curve, measured mean-square position, $\bar{y}^2$ (μm^2), versus time, t (s), after release. Lower curve, mean position $\bar{y}$ (μm). Error bars are smaller than the plot symbols. *c*, Measured probability distributions, at the same times after release as *b*, for motion perpendicular to the step edge. The vertical line indicates the starting position, $0.56\text{ }\mu\text{m}$. Inset, plot of the mean position, $\bar{x}$ (μm), versus t .



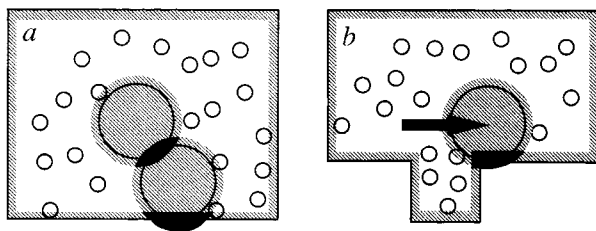


FIG. 3 Illustrations of the entropic interactions of hard spheres with surfaces. Owing to the hard-core repulsion, the small spheres' centres of mass are excluded from the hatched regions surrounding the large spheres and the walls. When the large spheres are separated from each other and from the walls, the volume accessible to the small spheres is the total volume of the box minus the hatched regions. *a*, When a large sphere touches another large sphere or a wall, the volume accessible to the small spheres (and, therefore, the entropy) increases by the volume of the 'excluded-volume overlap' region shown in black. This increase in entropy induces an attractive force between large spheres and walls. *b*, When a large sphere protrudes over a trench the excluded-volume overlap decreases, so that the sphere is repelled from the edge, as indicated by the arrow.

energy increased, equivalent to a force of $\sim 0.04 \times 10^{-12}$ N pushing the sphere back away from the edge. The probability distributions in Fig. 2c illustrate the effect of this barrier on particle motion. First, the distribution was asymmetric after 4.1 s, decreasing dramatically to the left of the edge because few of the spheres could 'escape' laterally across the edge of the terrace. Second, the peak of the probability distribution shifted from the starting point, towards the edge where the large spheres accumulated. If the large spheres simply diffused freely across the edge and disappeared from view, the probability-distribution peak would have shifted away from the edge. Third, the mean particle position, $\bar{x}$, increased (moved away from the edge) with time because the spheres were repelled from positions less than zero. We point out that the small dip in the free energy and the small peak in the late-time probability distribution near $-0.23 \mu\text{m}$ correspond to spheres that escaped across the barrier and were attracted to the vertical surface of the step. Last, the origin of the shallow free-energy minimum located $0.3 \mu\text{m}$ to the right of the edge is uncertain but may result from a non-uniformity in the small-sphere concentration induced by the step edge.

We stress that the lifetime of large spheres on the surface depends only on the concentration of small spheres. In a sample with small-sphere volume fraction reduced by 2 (to 0.15), the average lifetime of a large sphere on the surface was reduced by 3. Because other interactions (for example, electrostatic, van der Waals, or hydration forces) should not depend strongly on the small-sphere concentration, we concluded that they are much too weak to account for the forces here. (A measurement of the interaction between a polystyrene sphere and a glass wall is reported in ref. 22.)

To understand the origin of the repulsive force observed in the experiment, it is instructive to review the basic ideas behind entropically driven forces. A mixture of hard spheres maximizes its entropy by maximizing the volume accessible per particle. Although there exist only hard-core repulsions between pairs of particles, maximizing the entropy in the binary mixture can lead to an entropic attraction between the larger particles^{23,24} and between the particles and the walls^{16,25-27}. Figure 3a shows that, when a large sphere approaches another large sphere or the wall, the total volume available to the small spheres increases. This increases the total entropy of the mixture (decreases the Helmholtz free energy) by an amount proportional to the size of the 'excluded-volume overlap' region (represented in black) multiplied by the pressure of the small spheres. Specifically, moving a single large sphere from the bulk to the surface decreases the mixture's Helmholtz free energy by approximately $3(a_L/a_S)\phi_S k_B T$,

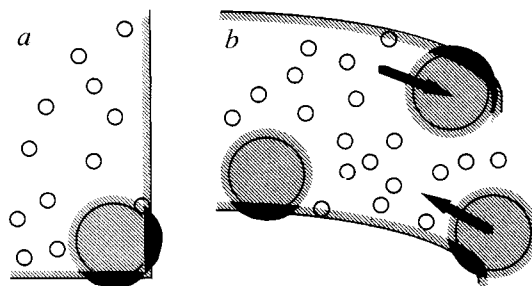


FIG. 4 Illustration of surface geometries that can trap the larger particles or induce drift. *a*, In a corner, the overlap volume (black region) is approximately twice as large as on a flat wall. Large spheres would tend to remain here for several seconds or minutes. Furthermore, the dense row of spheres in the corner could induce nucleation of highly ordered, three-dimensional crystals. *b*, Along a wall of changing curvature, the excluded-volume overlap changes with position. The larger sphere is driven along the wall in the direction of increasing overlap, as indicated by the arrows. Drift speeds up to approximately $0.1a_L$ per second may be obtained, where a_L is the diameter of the larger spheres.

where $a_L(a_S)$ is the large- (small-) sphere diameter, $\phi_S = N_S \pi a_S^3 / (6V)$, and N_S is the number of small spheres in the sample of volume V (ref. 16). This free-energy gradient is equivalent to a force of $\sim 10^{-12}$ N pushing the large sphere towards the wall.

Consider a large sphere attracted by the depletion force to a surface into which a trench of width w has been cut (Fig. 3b). As the large sphere moves along the surface towards the trench, the amount of excluded-volume overlap changes. If the trench is wide enough to admit small spheres ($w > a_S$), then the excluded-volume overlap decreases (the free energy increases) as the large sphere protrudes over the edge of the trench; the trench acts like a barrier against the large sphere. The step edge of our experiment may be thought of as a trench with infinite w .

For comparison to our results, we have calculated the Helmholtz free energy, F , of the mixture as a function of the distance of a single $0.46 \mu\text{m}$ sphere from the edge, considering only the entropic excluded-volume effect. At the edge, there is a substantial free-energy barrier, the magnitude of which depends on the trajectory assumed for the large sphere. If we assume that it clings to the surface as it moves down around the corner of the step edge, the free energy increases to approximately $3k_B T$ at the edge, then returns to zero at $-0.23 \mu\text{m}$, where the sphere lies on the vertical surface (solid curve, Fig. 2a). By contrast, if we assume that the sphere moves at constant height, starting on the terrace far from the edge and moving horizontally past the edge, the free energy barrier exceeds $5k_B T$ (not shown). When the large sphere is slightly above the terrace, the magnitude of the barrier is reduced because the excluded-volume overlap is smaller. The fact that some of the particles in the experiment were slightly above the surface could explain why the measured barrier is less than $3k_B T$. The fact that the measured free energy does not return to zero to the left of the step at $-0.23 \mu\text{m}$ simply reflects our observation that some of the particles escaped into the bulk (where $F \approx 5k_B T$) while others went to the vertical surface of the step (where $F \equiv 0$).

Building on the effects demonstrated here, it is possible to devise structures that create potentially useful, localized and directional entropic force fields. A microscopic circular terrace could be used to trap a single large sphere which would be attracted vertically to the flat surface and constrained laterally by the step-edge barrier. An array of such terraces could be used to make the large spheres self-assemble in a chosen pattern. In Fig. 4a, we illustrate a large sphere attracted to an inside corner. In Fig. 4b, we show how net drift of large spheres over a large distance may be induced along a surface with changing curvature.

The strength and range of these entropic force fields can be controlled by varying ϕ_s , a_L and a_s , while the location and direction are determined by the substrate geometry. In practical systems, the effects are several $k_B T$ in magnitude and tens or hundreds of nanometres in range. □

Received 29 April; accepted 9 August 1996.

1. Pusey, P. N. & van Megen, W. *Nature* **320**, 340–342 (1986).
2. Dinsmore, A. D., Yodanis, A. G. & Pine, D. J. *Phys. Rev. E* **52**, 4045–4057 (1995).
3. Bartlett, P., Ottewill, R. H. & Pusey, P. N. *Phys. Rev. Lett.* **68**, 3801–3804 (1992).
4. Sanyal, S., Easwar, N., Ramaswamy, S. & Sood, A. K. *Europhys. Lett.* **18**, 107–110 (1992).
5. van Duljneveldt, J. S., Heinen, A. W. & Lekkerkerker, H. N. W. *Europhys. Lett.* **21**, 369–374 (1993).
6. Imhof, A. & Dhont, J. K. G. *Phys. Rev. Lett.* **75**, 1662–1665 (1995).
7. Yasrebi, M., Shih, W. Y. & Aksay, I. A. *J. Colloid Interface Sci.* **142**, 357–368 (1991).
8. Steiner, U., Meller, A. & Stavans, J. *Phys. Rev. Lett.* **74**, 4750–4753 (1995).
9. Bibette, J., Roux, D. & Pouligny, B. *J. Phys. II Fr.* **2**, 401–424 (1992).
10. Calderon, F. L., Blais, J. & Bibette, J. *Colloids Surf. A* **74**, 303–309 (1993).
11. De Hek, H. A. & Vrij, A. *J. Colloid Interface Sci.* **84**, 409–422 (1981).
12. Calderon, F. L., Bibette, J. & Blais, J. *Europhys. Lett.* **23**, 653–659 (1993).
13. Ilett, S. M., Orrock, A., Poon, W. C. K. & Pusey, P. N. *Phys. Rev. E* **51**, 1344–1352 (1995).
14. Herzfeld, J. *Acc. Chem. Res.* **29**, 31–37 (1996).
15. Buitenhuis, J., Donselaar, L. N., Buining, P. A., Stroobants, A. & Lekkerkerker, H. N. W. *J. Colloid Interface Sci.* **175**, 46–56 (1995).
16. Kaplan, P. D., Rouke, J. L., Yodanis, A. G. & Pine, D. J. *Phys. Rev. Lett.* **72**, 582–585 (1994).
17. Yablonovitch, E. *J. Opt. Soc. Am.* **10**, 283–295 (1993).
18. Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, 1995).
19. Murray, C. A. & Grier, D. G. *Am. Sci.* **83**, 238–245 (1995).
20. Bangs, L. B. *J. Clin. Immunology* **13**, 127–131 (1990).
21. Crocker, J. C. & Grier, D. G. *Phys. Rev. Lett.* **73**, 352–355 (1994); *J. Colloid Interface Sci.* **179**, 298–310 (1996).
22. Prieve, D. C. & Frej, N. A. *Langmuir* **6**, 396–403 (1990).
23. Asakura, S. & Oosawa, F. *J. Polym. Sci.* **33**, 183–192 (1958).
24. Vrij, A. *Pure Appl. Chem.* **48**, 471–483 (1976).
25. Kaplan, P. D., Fauchaux, L. P. & Libhaber, A. J. *Phys. Rev. Lett.* **73**, 2793–2796 (1994).
26. Sober, D. L. & Walz, J. Y. *Langmuir* **11**, 2352–2356 (1995).
27. Milling, A. & Biggs, S. J. *Colloid Interface Sci.* **170**, 604–606 (1995).
28. Ashkin, A., Dziedzic, J. M., Bjorkholm, J. E. & Chu, S. *Opt. Lett.* **11**, 288–290 (1986).
29. Svoboda, K. & Block, S. M. *Annu. Rev. Biophys. Biomolec. Struct.* **23**, 247–285 (1994).

ACKNOWLEDGEMENTS. We thank T. C. Lubensky and P. D. Kaplan for useful conversations, and A. T. Johnson, Z. M. Yu, T. Heinzel and G. J. Dolan for their help in silicon processing. This work was supported by the US NSF and the PENN Materials Research Laboratory.

CORRESPONDENCE should be addressed to A.D.D. (e-mail: dinsmore@sol1.lirm.upenn.edu).

Direct U–Th dating of marine sediments from the two most recent interglacial periods

Niall C. Slowey*, Gideon M. Henderson† & William B. Curry‡

* Department of Oceanography, Texas A&M University, College Station, Texas 77843, USA

† Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

‡ Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

A KNOWLEDGE of the age of marine sediments is necessary to determine the timing of events and rates of processes in the marine realm, and the relationships among marine and other climatically sensitive records. The establishment of an accurate chronology for Pleistocene marine sediments beyond the range of radiocarbon dating (approximately the past 45 kyr) has therefore been a goal of palaeoceanographers for decades. Early attempts^{1,2} based on measurements of the radionuclides ²³⁰Th and ²³¹Pa were beset with problems, and subsequent studies focused on tying fluctuations in marine sediment oxygen-isotope records to events such as the formation of coral reef terraces and changes in the Earth's magnetic polarity^{3,4}, and tuning the resultant chronologies to the Earth's orbitally driven insolation variations^{5–8}. But

these chronologies (especially the age and duration of the last interglacial period) have been challenged by several studies^{9–12}, raising questions about the fundamental cause of Pleistocene climate fluctuations. Here we report the direct U–Th dating of aragonite-rich marine sediments from the Bahamas, and present an accurately dated marine oxygen-isotope record for the last two interglacials. We obtain dates of 120–127 kyr BP for the last interglacial and 189–190 kyr BP for the late stage 7 interglacial. These dates are in accord with the general theory of orbitally forced climate fluctuations and demonstrate the potential of our direct-dating approach for developing an absolute chronology for the Pleistocene marine oxygen-isotope record.

Algae and inorganic processes on the shallow edges and tops of Little Bahama Bank produce fine-grained aragonite with uranium concentrations far in excess of those found in foraminifera^{13–15} and which equal or exceed those of corals^{16,17}. Tides and storms cause this aragonite to become suspended in the water and transport it off the bank, whereupon it settles to the sea floor together with the normal rain of pelagic material¹⁸. Subaerial exposure of the banktop restricts aragonite production to the edge of the bank during sea-level lowstands, but the banktop sheds large quantities of material when flooded during interglacial highstands (sea level <15 m below present). Consequently, slope sediments possess expanded interglacial sections^{19–23} of U-rich marine sediments²⁴. They contain little terrigenous material, so detrital U and Th contents are small, and the shallow water depths mean that concentrations of Th scavenged from the water column are low. As these sediments have always been bathed in sea water they have not experienced diagenesis associated with exposure to meteoric water. The presence of typical marine microfossils and high primary U concentrations, coupled with low concentrations of detrital and scavenged Th, make interglacial Bahamian slope sediments ideal for direct U–Th dating of marine $\delta^{18}\text{O}$ record.

We sampled two wide-diameter piston cores (152JPC and 149JPC) from the southwestern slope of the Little Bahama Bank to test this approach (Table 1, Fig. 1). The cores were collected from the crests of ridges which extend down the slope, in order to isolate the sediment from possible disturbance by down-slope processes. In 152JPC, downcore variations of $\delta^{18}\text{O}$ and carbonate mineralogy indicate that sediments from the Holocene epoch and the last glaciation are not present. They slumped away

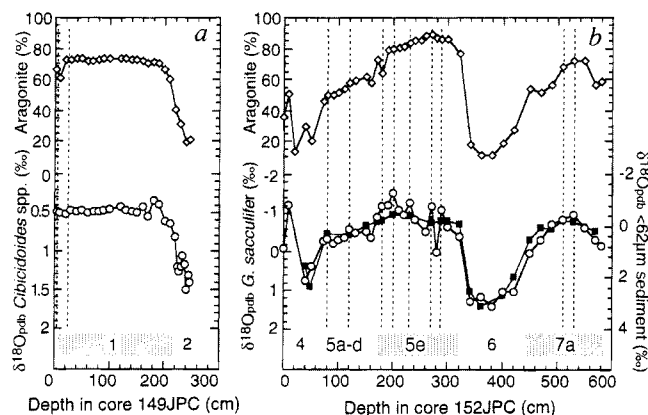


FIG. 1 The bulk-sediment aragonite percentage (diamonds), foraminiferal $\delta^{18}\text{O}$ (circles), and <62- μm sediment fraction $\delta^{18}\text{O}$ records (filled squares) of core 149JPC (refs 23, 25; a) and core 152JPC (b). Vertical dashed lines indicate stratigraphic positions where samples were taken for radiocarbon and U–Th dating. Aragonite content was measured by X-ray diffraction. Foraminiferal $\delta^{18}\text{O}$ values are averages of two or more replicate analyses; <62- μm sediment $\delta^{18}\text{O}$ values are typically single analyses (analytical uncertainty of any individual $\delta^{18}\text{O}$ analysis is 0.08‰). Marine $\delta^{18}\text{O}$ stages, as determined from the foraminiferal $\delta^{18}\text{O}$ stratigraphies, are shown by the shaded bars.

TABLE 1 Measurements on the <62- μ m fraction of Bahamian marine sediments

Sample (cm)	$^{238}\text{U}^*$ (p.p.m.)	$^{230}\text{Th}^*$ (pg g $^{-1}$)	$^{232}\text{Th}/^{230}\text{Th}^*$	$\delta^{234}\text{U}(\text{O})_{\text{raw}}^{*\dagger}$	$^{230}\text{Th}/^{238}\text{U}^{\ddagger}$ (activity ratio)	Age $_{\text{raw}}^{\dagger}$ (kyr)	CaCO $_3$ § (%)	Al (p.p.m.)	Detrital (%)	$\delta^{234}\text{U}(\text{T})_{\text{cor}}^{*\star}$	Age $_{\text{cor}}^{*\star}$ (kyr)
OC205-2-149JPC (26.2595° N, 77.6715° W, 423 m water depth)											
5	3.578 (6)	12.2 (1)	16,160 (165)	153.5 (3.2)	0.2089 (0.0009)	21.6 (0.2)	95.1	970	0.85	162 (6)	3.9 (8.9)
25	4.150 (8)	15.7 (2)	13,640 (231)	151.6 (3.7)	0.2329 (0.0014)	24.4 (0.3)	96.5	1291	1.21	166 (7)	12.4 (8.3)
OC205-2-152JPC (26.2267° N, 77.6708° W, 577 m water depth)											
80	2.803 (8)	30.5 (4)	8,646 (96)	124.8 (3.0)	0.6679 (0.0048)	95.7 (2.2)	98.4	900	0.87	164 (8)	71.6 (10.3)
120	5.565 (4)	63.3 (5)	3,477 (107)	101.1 (2.0)	0.6979 (0.0025)	106.8 (1.3)	96.5	671	0.58	136 (4)	93.4 (5.9)
180	3.955 (3)	51.4 (4)	2,782 (21)	112.9 (2.2)	0.7981 (0.0027)	132.6 (1.8)	96.8	677	0.60	165 (5)	120.8 (5.8)
200	4.839 (5)	62.2 (6)	1,027 (9)	111.1 (2.1)	0.7894 (0.0040)	130.3 (2.6)	96.2	262	0.16	160 (3)	124.6 (3.7)
230	4.914 (15)	63.3 (2)	940 (2)	117.6 (3.1)	0.7906 (0.0016)	129.0 (1.2)	97.0	274	0.20	169 (5)	124.5 (2.4)
270a	4.928 (13)	62.5 (2)	1,405 (5)	107.3 (3.6)	0.7780 (0.0016)	127.8 (1.3)	97.4	391	0.32	154 (6)	121.5 (3.2)
270b	5.147 (4)	65.2 (4)	1,452 (7)	105.5 (1.8)	0.7780 (0.0023)	128.2 (1.5)				151 (3)	121.4 (3.5)
287a	5.465 (11)	71.5 (2)	2,300 (6)	105.1 (3.2)	0.8036 (0.0016)	136.5 (1.4)	97.3	820	0.75	156 (6)	127.0 (4.8)
287b	5.152 (7)	66.7 (7)	2,433 (11)	104.4 (3.7)	0.7945 (0.0025)	133.7 (1.9)				154 (6)	124.1 (5.1)
510	4.859 (9)	72.9 (3)	1,266 (4)	78.2 (4.3)	0.9206 (0.0021)	198.5 (3.7)	97.8	297	0.24	137 (8)	190.3 (5.0)
530	4.728 (18)	71.3 (3)	1,490 (6)	82.6 (3.7)	0.9255 (0.0027)	196.9 (4.0)	97.7	320	0.26	144 (7)	190.0 (5.7)

In the first column, 'a' and 'b' are analyses of splits of the well mixed sediment treated independently through the analytical procedure. Numbers in parentheses throughout the table represent 2σ uncertainties on the last digit.

* Measured by thermal ionization mass spectrometry using methods similar to ref. 39. Accuracy was assessed by repeated measurement of the U500 standard which yielded a $^{235}\text{U}/^{238}\text{U}$ ratio of 95.85 ($n = 10$; $s.d. = 0.2$): within error of the certified value. The U blank was 100 ± 40 pg ($n = 4$) and is insignificant. The Th blank was 700 ± 300 pg ($n = 4$) and is corrected for and the errors propagated.

† Raw ages and $\delta^{234}\text{U}$ values are calculated directly from the measured U and Th values. $\delta^{234}\text{U}$ in per mil units equals $[(^{234}\text{U}/^{238}\text{U})_{\text{sample}} - (^{234}\text{U}/^{238}\text{U})_{\text{secular equilibrium}}] \times 1,000$ where $^{234}\text{U}/^{238}\text{U}$ is the atomic ratio. $\delta^{234}\text{U}(\text{O})$ is calculated using measured values and $\delta^{234}\text{U}(\text{T})$ is the initial value at the time of sediment deposition calculated using the sample's age and the uranium isotope decay constants.

‡ Values are calculated from measured ^{238}U and ^{230}Th concentrations.

§ Values are the average of duplicate or triplicate determinations based upon the pressure of purified CO_2 gas produced by acidification of weighed samples. Analytical uncertainty of individual CaCO_3 values, assessed by comparing values from individual analyses of pure CaCO_3 standard with those predicted from a linear fit to standards, is 0.4%.

|| Values are the average of duplicate determinations by neutron activation analysis using methods like ref. 40 ($s.d.$ of duplicate analyses $\sim 5\%$).

$^{\parallel}$ Detrital % calculated using the CaCO_3 and Al concentrations assuming sediments composed of calcium carbonate, silica and detrital minerals with 11 ± 2 wt% Al and 29 wt% Si.

* Corrected $\delta^{234}\text{U}(\text{T})$ and ages are corrected for detrital hosted U and Th (using the detrital % and assuming 2.8 ± 0.4 p.p.m. U and 13.5 ± 3.0 Th) and for ^{230}Th scavenged from sea water assuming a $^{232}\text{Th}/^{230}\text{Th}$ ratio in shallow Bahamian sea water of $10,000 \pm 4,000$. Errors are propagated to include analytical uncertainty and uncertainty due to both the corrections.

or were not recovered by the corer (not uncommon with piston cores). However, the $\delta^{18}\text{O}$ and mineralogy stratigraphies match closely the sections of other cores from the region that correspond to $\delta^{18}\text{O}$ stages 5 to 7, so we believe sediments deposited during this interval are essentially undisturbed by downslope processes and were recovered intact. In 149JPC, foraminiferal $\delta^{18}\text{O}$, carbonate mineralogy^{23,25} and AMS ^{14}C dates on the <62- μ m-sized fraction (5-cm sample, $1,390 \pm 35$ yr; 25-cm sample: $2,370 \pm 25$ yr) indicate that the sediments are of Holocene age.

Interglacial sedimentation rates in these cores are in excess of 10 cm kyr^{-1} and the 2.2‰ difference in foraminiferal $\delta^{18}\text{O}$ between stages 6 and substage 5e is large, suggesting that bioturbation has not degraded the temporal resolution of the records. Moreover, any mixing that might occur with older sediments will not introduce significant amounts of old aragonite into interglacial sections, as the glacial sections have less than 15% aragonite (Fig. 1). Interglacial sediments are almost pure carbonate (95–99%) and their aragonite contents are high (for example, 77–89% in substage 5e and 68–72% in substage 7a). Production of banktop sediments decreases during the lowstands, so the sediment section corresponding to stage 6 was deposited at a much lower rate and those corresponding to substages 5b and 5d are condensed and not readily apparent. The low sedimentation rates and, especially, the low aragonite contents of the glacial sediments make them less suitable for U–Th dating.

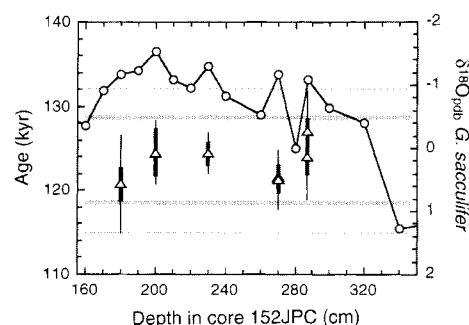
Figure 1 shows where sediments in cores 152JPC and 149JPC were sampled for U–Th dating. Samples were washed through 62- μ m sieves to separate bank-derived aragonite (primarily silt- and clay-sized particles^{17,20}) from U-poor sand-sized components like foraminifera, pteropods and sponge spicules. Table 1 details the analyses performed on this <62- μ m fraction. Though the <62- μ m fraction has a greater $\delta^{18}\text{O}$ range than *Globigerinoides sacculifer* (because the composition of the fine fraction varies with depth in the core), it is important that the downcore profiles of these two

records are quite similar (Fig. 1b). The comparison supports the notion that the bank-derived aragonite and foraminifera are the same age. It also suggests that a marine $\delta^{18}\text{O}$ stratigraphy for 152JPC could have been generated from the <62- μ m fraction alone, which is consistent with previous work showing bulk sediment can preserve high-quality marine $\delta^{18}\text{O}$ records²⁶.

Raw dates were calculated using the measured isotope ratios and concentrations (Table 1). These dates must be corrected for the presence of U and Th in detrital particles and for Th scavenged from sea water (reasonable corrections result in younger dates). The corrections for detrital U and Th were made using the measured Al concentrations of the samples and by assuming typical shale compositions²⁷ for U, Si and Th. Calculated clay concentrations are always less than 1.2% and the detrital correction typically alters the dates by less than their analytical uncertainty. Measured $^{232}\text{Th}/^{230}\text{Th}$ ratios are also corrected for detrital Th in this manner. ^{232}Th not accounted for by the detrital correction is assumed to be scavenged from sea water so, together with the $^{232}\text{Th}/^{230}\text{Th}$ ratio of sea water, detrital-corrected $^{232}\text{Th}/^{230}\text{Th}$ ratios can be used to correct the dates for scavenged ^{230}Th . The 5-cm sample from 149JPC is very young and has insignificant ingrown ^{230}Th from U decay, so its detrital-corrected $^{232}\text{Th}/^{230}\text{Th}$ ratio of 8,027 represents the sea water value for this ratio. This value is close to that of direct measurements on filtered sea water from the Sargasso Sea (J. Hoff, R. Edwards, K. Buesseler and R. Belackstock, personal communication). We assumed a $^{232}\text{Th}/^{230}\text{Th}$ ratio for scavenged Th of $10,000 \pm 4,000$ (back-calculated for ^{230}Th decay for old samples) which allows for variations in the value over time. Table 1 shows corrected dates with errors propagated to include the uncertainty in the two corrections and the analytical uncertainty.

We find no evidence that our dates are significantly affected by diagenesis. Uranium concentrations for all interglacial samples (2.8–5.5 p.p.m.) are in agreement with reported values for algal

FIG. 2 Expanded plot of age and $\delta^{18}\text{O}$ versus depth for substage 5e in core 152JPC. Substage 5e is identified by the extreme values of $\delta^{18}\text{O}$ (circles). The U–Th dates (triangles) are plotted with their analytical uncertainty (thick error bars, 2σ error or raw age in Table 1) and with the total of their analytical and correction uncertainties (thin error bars, 2σ error of corrected age in Table 1). Note that the duration of substage 5e is defined by the two samples at 180 and 287 cm. As these samples have very similar $^{232}\text{Th}/^{230}\text{Th}$ ratios and clay fractions, the magnitude of their age corrections is also very similar. Thus error associated with the correction affects these two ages by the same magnitude and does not significantly alter the duration of substage 5e implied by these data. The thick stippled lines therefore represents our best estimate of the maximum possible duration of 5e as ~ 10 kyr. The thin stippled lines represent a more conservative estimate which includes the error in the corrections and yields a maximum duration of ~ 17 kyr.



and inorganically precipitated aragonite^{16,17}. The ^{14}C and U–Th ages of Holocene sediments in core 149JPC are in broad agreement. Corrected $\delta^{234}\text{U}(T)$ values of all samples average 155% and range from 136 to 169% . They are generally slightly greater than the value ($\sim 148 \pm 4\%$) of modern coralline aragonite^{12,28–31}, but they do not increase with age as has been observed with corals affected by progressive diagenesis^{28,30,31}. The grouping of the substage 5e dates also suggests diagenesis is not a problem and this is supported by our two substage-7a dates of ~ 190 kyr that, given their uncertainty, agree with one another and with dates from unaltered corals²⁸.

Based on these considerations, we conclude that the corrected dates of samples from the last two interglacials are reliable and provide the first direct dates for these portions of the marine $\delta^{18}\text{O}$ record. As discussed above, the dates for substage 7a agree with those of unaltered corals; they are also consistent with the Devils Hole^{10,11} and SPECMAP⁷ chronologies. The date from 80 cm in core 152JPC indicates the transition from stage 5 to stage 4 began ~ 72 kyr ago and, within uncertainty, agrees with speleothem and well-preserved coral dates^{32,33}, and the Devils Hole^{10,11} and SPECMAP⁷ chronologies.

The seven dates from the one-metre-thick section of oxygen-isotope substage 5e range from 120 to 127 kyr and average 123.5 ± 4.5 yr (2 standard deviations) (Fig. 2). Dates of samples with the lowest $^{232}\text{Th}/^{230}\text{Th}$ ratios (200 and 230 cm) are particularly robust because the dates cannot be changed by more than 1 or 2 kyr using any reasonable age correction. Other substage-5e samples give dates broadly consistent with these two samples despite different magnitudes of correction. Considering analytical uncertainty, the maximum duration of the last interglacial implied by these dates is ~ 10 kyr (Fig. 2). The actual duration of substage 5e could be slightly longer (~ 1 – 2 kyr) because no date exists for

the very earliest part of the substage 5e section. It is worth noting that the magnitudes of corrections for the youngest and oldest samples (180 and 287 cm) are within 1 kyr, reflecting their rather similar $^{232}\text{Th}/^{230}\text{Th}$ ratios and clay fractions. Therefore, any consistent error associated with the corrections has a negligible effect on the difference between the ages of these samples and the estimate of the duration of substage 5e.

These U–Th dates represent unequivocal dates for the whole substage-5e interval of the marine $\delta^{18}\text{O}$ record. Their distribution is consistent with summarized distributions of dates (~ 122 – 127 kyr) from highstand corals that are judged to be reliable based on their $\delta^{234}\text{U}$ values^{12,29,34}. In comparison, the event equated with substage 5e in the Devils Hole record has an earlier onset and longer duration (~ 140 and ~ 20 kyr, respectively)^{10,11} than substage 5e as defined by the marine record. It has been suggested that the regional hydrological cycle in the southwestern United States influences the timing of changes in the Devils Hole record^{35,36}. Our dates imply a timing and duration for substage 5e in substantial agreement with the orbitally tuned marine chronology⁷. Uncertainty in the dates allows for the possibility of a slightly earlier onset to the last interglacial than indicated by the SPECMAP chronology, but such an onset would still be consistent with orbital forcing³⁷.

Initial direct U–Th dating of the marine $\delta^{18}\text{O}$ record supports the theory that variations in the Earth's orbit are a fundamental cause of Pleistocene climate changes. Further application and refinement of the absolute dating approach utilized here should provide additional insight into the mechanisms driving Pleistocene climate fluctuations. For example, we should now be able to determine more precisely the absolute relationships between marine records, orbital forcing and ice-age CO_2 changes (see, for example, ref. 38). □

Received 20 February; accepted 6 August 1996.

- Sackett, W. M. *Science* **132**, 1761–1762 (1960).
- Rosholt, J. N., Emiliani, C., Geiss, J., Koczy, F. F. & Wangersky, P. J. *J. Geol.* **69**, 162–185 (1961).
- Mesolella, K. J., Matthews, R. K., Broecker, W. S. & Thurber, D. L. *J. Geol.* **77**, 250–274 (1969).
- Shackleton, N. J. & Opdyke, N. D. *Quat. Res.* **3**, 39–55 (1973).
- Hays, J. D., Imbrie, J. & Shackleton, N. J. *Science* **194**, 1121–1132 (1976).
- Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A. et al.) 269–305 (Reidel, Dordrecht, 1984).
- Martinson, D. G. et al. *Quat. Res.* **27**, 1–29 (1987).
- Shackleton, N. J., Berger, A. & Peltier, W. R. *Trans. Roy. Soc. Edinburgh* **81**, 251–261 (1990).
- Jouzel, J. et al. *Nature* **364**, 407–412 (1993).
- Ludwig, K. R. et al. *Science* **258**, 284–287 (1992).
- Winograd, I. J. et al. *Science* **255**, 255–260 (1992).
- Szabo, B. J., Ludwig, K. R., Muhs, D. R. & Simmons, K. R. *Science* **266**, 93–96 (1994).
- Delaney, M. L. & Boyle, E. A. *Earth Planet. Sci. Lett.* **62**, 258–262 (1983).
- Russell, A. D., Emberson, S., Nelson, B. K., Erez, J. & Lea, D. W. *Geochim. Cosmochim. Acta* **58**, 671–681 (1994).
- Henderson, G. M. & O'Nions, R. K. *Geochim. Cosmochim. Acta* **59**, 4685–4695 (1995).
- Sackett, W. M., Mo, T., Spalding, R. F. & Exner, M. E. in *Radioactive Contamination of the Marine Environment* 757–769 (Int. Atomic Energy Agency, Vienna, 1993).
- Milliman, J. D. *Marine Carbonates* (Springer, Berlin, 1974).
- Neumann, A. C. & Land, L. S. *J. Sedim. Petrol.* **45**, 763–786 (1975).
- Droxler, A. W., Schlager, W. & Wallon, C. C. *Geology* **11**, 235–239 (1983).
- Boardman, M. R. & Neumann, A. C. *J. Sedim. Petrol.* **54**, 1110–1123 (1984).
- Boardman, M. R. et al. *Geology* **14**, 28–31 (1986).
- Burns, S. J. & Neumann, A. C. *Mar. Geol.* **77**, 277–286 (1987).
- Slowey, N. C. & Curry, W. B. *Nature* **358**, 665–668 (1992).
- Slowey, N. C. & Curry, W. B. *Paleoceanography* **6**, 609–619 (1991).

- Slowey, N. C. & Curry, W. B. *Paleoceanography* **10**, 715–732 (1995).
- Shackleton, N. J., Hall, M. A., Pate, D., Meynadier, L. & Valet, P. *Paleoceanography* **8**, 141–148 (1993).
- Taylor, S. R. & McClelland, S. M. *The Continental Crust: Its Composition and Evolution* (Blackwell Scientific, Oxford, 1985).
- Gallup, C. D., Edwards, R. L. & Johnson, R. G. *Science* **263**, 796–800 (1994).
- Stirling, C. H., Esat, T. M., McCulloch, M. T. & Lambeck, K. *Earth Planet. Sci. Lett.* **135**, 115–130 (1995).
- Henderson, G. M., Cohen, A. S. & O'Nions, R. K. *Earth Planet. Sci. Lett.* **115**, 65–73 (1993).
- Bard, E., Fairbanks, R. G., Hamelin, B., Zindler, A. & Hoang, C. T. *Geochim. Cosmochim. Acta* **55**, 2385–2390 (1991).
- Richards, D. A., Smart, P. L. & Edwards, R. L. *Nature* **367**, 357–360 (1994).
- Bard, E., Hamelin, B. & Fairbanks, R. G. *Nature* **346**, 456–458 (1990).
- Kaufman, A. *Quat. Res.* **25**, 55–62 (1986).
- Imbrie, J. L., Mix, A. C. & Martinson, D. G. *Nature* **363**, 531–533 (1993).
- Crowley, T. J. *Paleoceanography* **9**, 1–5 (1994).
- Crowley, T. J. & Kim, K.-Y. *Science* **265**, 1566–1568 (1994).
- Sowers, T. et al. *Paleoceanography* **8**, 737–766 (1993).
- Edwards, R. L., Chen, J. H. & Wasserburg, G. J. *Earth Planet. Sci. Lett.* **81**, 175–192 (1986/1987).
- James, W. D. & Acevedo, L. E. *J. Radiol. Nucl. Chem. Articles* **171**, 287–302 (1993).

ACKNOWLEDGEMENTS. We thank P. Boothe, W. D. James, M. Raulerson and M. Yeager for assistance with analyses and discussions; R. Schneider and the National Ocean Sciences AMS Facility for radiocarbon analyses; R. Anderson, T. Crowley, M. Fleisher and N. Shackleton for discussions; and the reviewers for their comments. N.S. thanks C. Stearns for inspiration. This research and core curation were supported by the NSF, the LDEO and TAMU.

CORRESPONDENCE should be addressed to N.C.S. (e-mail: Slowey@ocean.tamu.edu).

Interaction of weak faults and non-newtonian rheology produces plate tectonics in a 3D model of mantle flow

Shijie Zhong & Michael Gurnis

Seismological Laboratory, California Institute of Technology, Pasadena, California 91125, USA

ACCORDING to the theory of plate tectonics, relatively rigid plates are bounded by large faults; plate motion has negligible internal strain^{1,2}, with significant toroidal component to the velocity³. By contrast, models of mantle flow with viscous rheology in an intact medium predict little toroidal component and substantial internal strain in surface motion⁴. It has been suggested⁵ that the observed characteristics of plate motion are related to faulted plate margins, which are observed to be weak⁶. Here we confirm this suggestion, using three-dimensional models of mantle flow that incorporate faults and the forces exerted on plates by subducting slabs ('slab pull') and mid-ocean ridges ('ridge push'). Our models show that plate-like motion results from the interaction between weak faults and a strain-weakening power-law rheology. Weak transform faults tend to guide plate motion. This guiding effect and the decoupling that occurs at thrust faults may result in oblique subduction. Convergent margins are associated with realistic trench and fore-bulge topography. By simultaneously predicting surface kinematics, topography and gravity, the models achieve a useful degree of tectonic realism.

Our previous two-dimensional models have indicated that

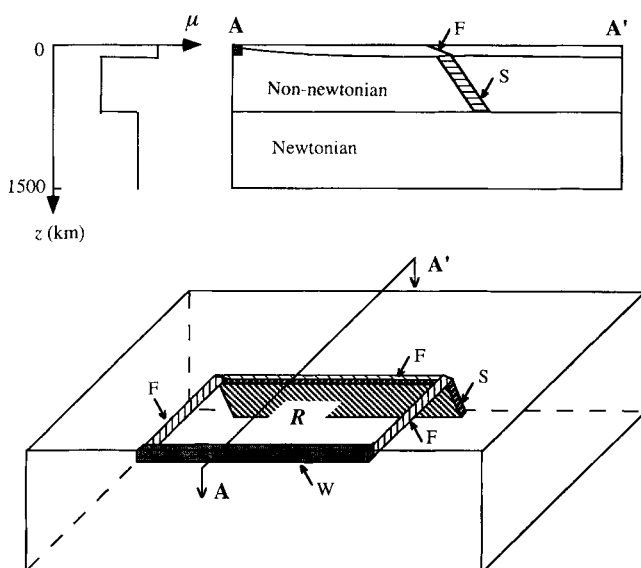


FIG. 1 Bottom, the geometry of the three-dimensional viscous flow model with faults. F, S and W stand for faults, slab and weak zone, respectively. Top left, the viscosities for the upper mantle and the top 100 km are the average μ_{eff} for these regions; the values of pre-exponent A (see text) are 1.5×10^7 , 1.5×10^3 and 15, for the top 100 km layer and slabs, spreading centres and the upper mantle, respectively. Viscosity is normalized by 2×10^{22} Pa s. Top right, a cross-section AA' of the model, showing viscosity and thermal structure. The medium overriding the mantle is water. The densities for water and mantle are $1,000 \text{ kg m}^{-3}$ and $3,300 \text{ kg m}^{-3}$, respectively. Thermal diffusivity and thermal expansivity are $10^{-6} \text{ m}^2 \text{ s}^{-1}$ and $2 \times 10^{-5} \text{ K}^{-1}$, respectively.

faulted convergent plate margins contribute to producing plate-like motion, and that the dynamically determined motion of faulted margins controls the evolution of plates and subduction dynamics^{7,8}. In this study the incorporation of three-dimensional faults allows us to study the influence of faults on poloidal and toroidal motion. Not only is fault generation a non-viscous phenomenon but, once formed, faults exist for long periods of geological time. Faults reflect the history-dependence of lithospheric deformation. Therefore, for instantaneous mantle-flow models, faults should be treated as pre-existing mechanical structures. In our models, faults are simulated as internal planes across which normal velocities are continuous but tangential velocities may be discontinuous (Fig. 1). Flow on either side of a fault can be coupled through frictional stress⁷. It must be emphasized that in our fault algorithm fault geometry is the only input and the sense of motion on faults (for example, thrust or strike slip) is determined by the dynamics.

Our three-dimensional cartesian models include a half-space cooling thermal boundary layer and cold subducted slabs (Fig. 1 shows one geometry). The buoyancy is derived as if a region, R, is spreading from the left side boundary at a velocity of 3 cm yr^{-1} (Fig. 1). In R, surface age increases with distance from the

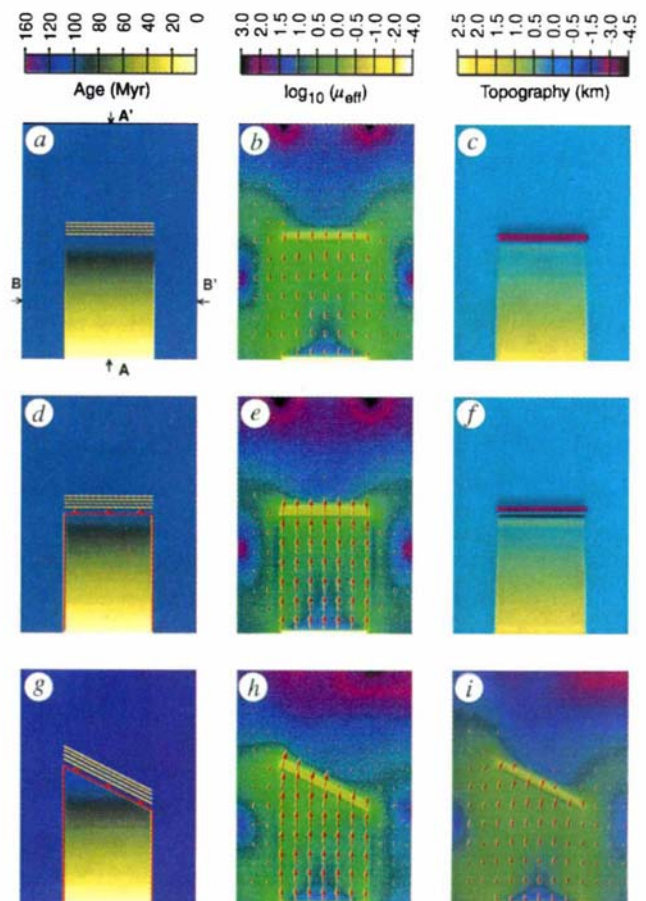


FIG. 2 The top (a, b and c) and middle (d, e and f) three panels are plan views of input buoyancy structure and fault geometry and output of surface velocity (red arrows), μ_{eff} and topography, for models NF and WF, respectively. Panels g and h show input parameters and output surface velocity and μ_{eff} for the non-parallel slab model with faults; panel i shows output surface velocity and μ_{eff} for the same model without faults. In a, d and g, yellow lines show the depth to slabs with 100, 300, 500 and 670 km depth contours; red lines over lithospheric age represent faults; triangles indicate the sense of fault dip. For these models, the length, width and thickness of the model boxes (in km) are 6,000, 4,500 and 1,500, respectively, which are covered by $64 \times 32 \times 16$ elements.

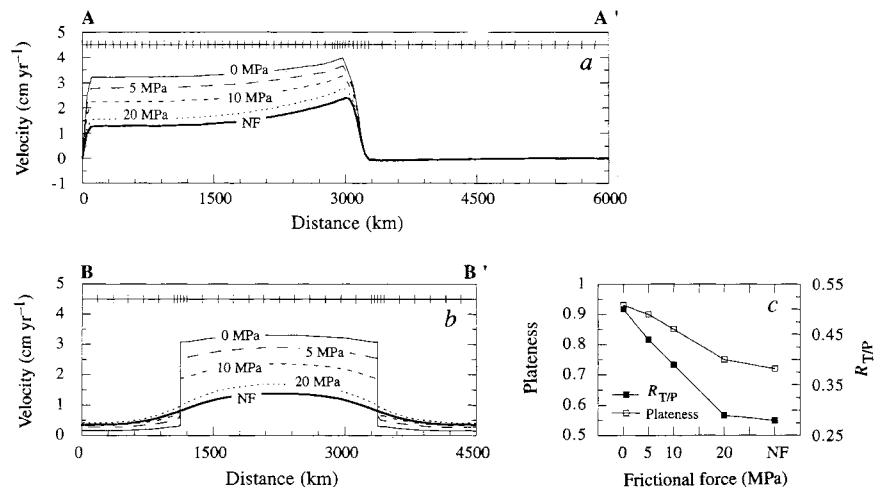


FIG. 3 Surface velocities for cross-sections AA' (a) and BB' (b) (see Fig. 2a for their positions), and $R_{T/P}$ and P (c) for models without faults (NF) and with faults of different frictional stress. In a and b, the unevenly distributed marks represent grid spacing.

spreading centre. Outside of R the thermal structure is identical to that with the oldest age. A 100-km-thick slab extends from 100 to 670 km depth with a 60° dip (Fig. 1). Slab buoyancy is identical to that within the lithosphere just before subduction. The viscosity in the lower mantle is newtonian (that is, 2×10^{22} Pa s). Viscosity above the 670 km depth is determined by a power-law rheology with $n = 3$ (that is, $\mu_{\text{eff}} = A^{1/n} \dot{\epsilon}^{1/n-1}$, where μ_{eff} is the effective viscosity, A is the pre-exponent related to temperature, n is an exponent and $\dot{\epsilon}$ is strain rate)⁹. At the spreading centre, A is decreased to account for the higher temperature and partial melting below a ridge. Values of A for the top 100 km layer and slabs, spreading centres, and upper mantle are chosen such that average μ_{eff} for the top 100 km and upper mantle are about 10^{23} Pa s and 4×10^{20} Pa s, respectively (Fig. 1). Except at the spreading centre, R is bounded by faults which dissect the top 100 km layer (Fig. 1). Fault dip is 30° when it is above slabs, but vertical elsewhere. The momentum and continuity equations are solved with free-slip bottom and top boundaries and reflecting conditions on side walls using a finite element code CITCOM¹⁰.

For comparison, the first model (NF) does not include faults, and slab strike is parallel to the spreading centre (Fig. 2a). The length and width of R are 3,000 km and 2,250 km, respectively. Surface velocities in R for this model vary gradually in both the spreading-parallel and perpendicular directions (Figs 2b and 3), indicating large internal strain. The average velocity in R along the spreading-parallel direction, V_p , is 1.5 cm yr⁻¹. The ratio of toroidal to poloidal components¹¹, $R_{T/P}$, is 0.28. The plateness, P , defined as $1 - V_{\text{rms}}/V_p$, is 0.72, where V_{rms} is the root-mean-square deviation of the velocity from V_p , in R , along the spreading-parallel direction. For a perfect plate with the geometry of R , $R_{T/P} = 0.55$ and $P = 1$; by comparison, for a sinusoidal variation in velocity in the spreading-parallel direction, $P = 0.52$. Both measures of the plate characterization have their limitations: P cannot account for the strain associated with velocity perpendicular to the spreading direction, and for certain geometries, $R_{T/P}$ can be insensitive to local strain. Accordingly, a model must have both large $R_{T/P}$ and large P to have plate-like characteristics. Both the divergent and convergent margins are weak (Fig. 2b); μ_{eff} in other regions varies gradually and reflects the variations in strain rate (Fig. 2b). Topography in R increases with age with a broad depression over the slab (~500 km wide) (Fig. 2c).

We now contrast model NF with a model (WF) that includes faults with zero frictional stress but is otherwise identical to model NF. The strike of the dipping fault is parallel to the spreading centre and perpendicular to that of vertical faults (Fig. 2d). Surface velocities in R bounded by faults are nearly constant ($P = 0.93$) and parallel to the strike of vertical faults with significant toroidal power ($R_{T/P} = 0.5$) (Figs 2e and 3). The sharp strike-slip motion across the vertical faults (Fig. 3) indicates that

the faults have the characteristics of transform faults. V_p is about 3.3 cm yr⁻¹, close to the velocity with which the buoyancy was derived. The gross topography and viscosity are similar to those of model NF (Fig. 2b, c, e and f). However, there is a prominent outer rise (~300 km wide and 0.4 km high) and a distinct trench (~100 km wide and 4 km deep) near the dipping fault (Fig. 2f). As in previous work⁷, the dynamic topography in the back-arc region (Fig. 2f) appears too deep compared with observations, but such a comparison is complicated by back-arc volcanism, which can both obscure the dynamic topography and (if uncompensated) affect the geoid. There is little strain near the vertical faults and because of the non-newtonian rheology the fluid surrounding the faults has a higher viscosity (Fig. 2e).

Frictional stress on faults and rheology have an important influence on surface motion. Models differing from model WF only in having a frictional force ranging from 5 to 20 MPa on faults show that as the frictional force increases, surface velocity, $R_{T/P}$ and P decrease and progressively resemble those with no faults (Fig. 3). A newtonian model that uses average μ_{eff} in each region from model WF (that is, lithosphere, weak spreading centre, slab, the upper and lower mantle) but is otherwise identical to model WF displays non-plate characteristics ($R_{T/P} = 0.15$; $P = 0.81$). In a model that includes a non-newtonian lithosphere but is otherwise identical to the preceding model, $R_{T/P}$ increases to 0.43. This suggests that besides weak faults, non-newtonian rheology in the lithosphere is essential to producing plate-like motion.

The strike of subducted slabs is not perpendicular to plate motion in many subduction zones where oblique subduction is often observed¹². The influence of the orientation of buoyancy force and weak faults on surface motion is studied with a model in which the strike of slab is non-parallel to the spreading centre (Fig. 2g). Surface velocities in R show little internal strain with a sharp contrast across the vertical faults (Fig. 2h) ($P = 0.91$ and $R_{T/P} = 0.49$; $R_{T/P} = 0.55$ for a perfect plate). Surface motion in R is predominantly parallel to the vertical faults and deviates significantly from the trench-normal direction near the dipping fault (Fig. 2d), consistent with the observation of oblique subduction. However, in an identical model but without the faults, surface velocities in R change their directions and become perpendicular to slab strike near the slab (Fig. 2i), showing substantial internal strain ($P = 0.73$; $V_p = 1.6$ cm yr⁻¹; $R_{T/P} = 0.30$). These two models indicate that the interaction between weak faults and non-newtonian rheology significantly influences the direction of plate motion and causes oblique subduction.

Plates usually have a complicated geometry with variable strikes of subduction zones. Around the Pacific plate, for example, the strike of Izu-Bonin subduction zone is approximately perpendicular to plate motion, whereas the Aleutian trench is nearly parallel to plate motion. A model in which R is bounded by

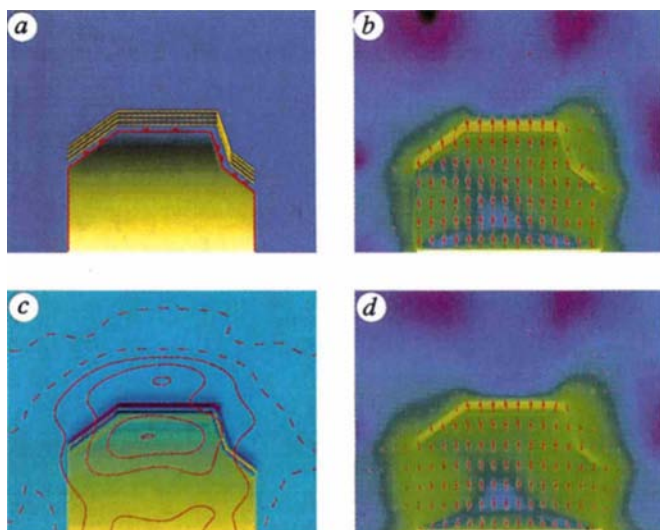


FIG. 4 Plan view of input buoyancy structure and faults (a), model output of surface velocity (red arrows) and μ_{eff} (b), topography and long-wavelength (>1,000 km) geoid (c) for the model with variable strikes of slabs, and surface velocity and μ_{eff} (d) for the model excluding the faults. Contour interval for geoid is 10 m and dashed lines are negative values. The colour scales are the same as in Fig. 2. The length, width and thickness of the model boxes (in km) are 6,000, 7,500 and 1,500, respectively, which are covered by $54 \times 54 \times 16$ elements.

faults and slabs with variable strikes (Fig. 4a) shows that surface velocities remain plate-like (Fig. 4b) with $P = 0.87$ and $R_{T/P} = 0.40$ ($R_{T/P} = 0.42$ for a perfect plate). Surface velocities in R are predominantly parallel to the vertical faults and show oblique subduction where the dipping faults are not parallel to the spreading centre (Fig. 4b). Trench topography is evident oceanward from the dipping faults and there is 40 m long-wavelength geoid high over slabs (Fig. 4c). Qualitatively, both trench/fore-bulge topography and the geoid are essentially identical to observed features in the western Pacific (for example, approximately 4 km deep and 150 km wide trenches, and 30 m long-wavelength geoid high). The reduced P results from the velocity variations in the spreading perpendicular direction near the short vertical fault (Fig. 4b) where slabs are younger; a more realistic temperature-dependent rheology may lead to a higher plateness. When the faults are excluded from the model, surface velocities vary gradually in both the spreading parallel and perpendicular directions with P decreased to 0.75 (Fig. 4d). However, $R_{T/P}$ only decreases slightly to 0.39, suggesting that a large fraction of toroidal power is not a sufficient criterion for characterizing the existence of plates.

These models show that the characteristics of plate motion can be explained with realistic mantle buoyancy, weak faults and a power-law rheology ($n = 3$). Besides plate kinematics, the models also simultaneously predict realistic trench/fore-bulge topography and the long-wavelength geoid (Fig. 4b and c). In the past, different physical processes were used to explain these observed features (for example, elastic plate bending models for trench/fore-bulge topography and laterally homogenous viscous flow models for the geoid). By simultaneously matching observed trench topography, plate kinematics and the long-wavelength geoid, our models will provide more realistic constraints on materials properties and slab dynamics. Weak transform faults enhance the motion along the faults. Weak transform faults and their strong adjacent regions cause transform faults to guide plate motion (Fig. 2e and h), consistent with inferences based on plate kinematics¹³. The guiding influence of transform faults on plate motion and the decoupling of thrust faults result in oblique subduction where the strike of thrust faults is non-perpendicular

to the transform faults (Figs 2h and 4b), consistent with observations^{12,14}. Non-zero normal velocities on faults indicate migrating plate margins. Ultimately, with the formulation presented here and with multiple plates coupled with the solution of energy equation, we should be able to predict the evolution of plate geometry as accomplished in two dimensions⁸. □

Received 12 January; accepted 6 August 1996.

1. Wilson, J. T. *Nature* **207**, 343–347 (1965).
2. Morgan, J. W. *J. Geophys. Res.* **73**, 1959–1982 (1968).
3. Hager, B. H. & O'Connell, R. J. *J. Geophys. Res.* **84**, 1031–1048 (1979).
4. Christensen, U. R. & Harder, H. *Geophys. J. Int.* **104**, 213–226 (1991).
5. Hager, B. H. & O'Connell, R. J. *J. Geophys. Res.* **86**, 4843–4867 (1981).
6. Lachenbruch, A. H. & Sass, J. H. *Geophys. Res. Lett.* **15**, 981–984 (1988).
7. Zhong, S. & Gurnis, M. J. *Geophys. Res.* **99**, 15683–15695 (1994).
8. Zhong, S. & Gurnis, M. *Science* **267**, 838–843 (1995).
9. Ranalli, G. *Rheology of the Earth* 326–336 (Chapman & Hall, London, 1995).
10. Moresi, L. N. & Gurnis, M. *Earth Planet. Sci. Lett.* **138**, 15–28 (1996).
11. Gable, C. W., O'Connell, R. J. & Travis, B. J. *J. Geophys. Res.* **96**, 8391–8405 (1991).
12. Jarrard, R. D. *Rev. Geophys.* **24**, 217–284 (1986).
13. Richards, M. A. & Engebreson, D. C. (abstr.) *EOS (spring suppl.)* **75**, 63 (1994).
14. McCaffrey, R. *Pur. Appl. Geophys.* **142**, 173–224 (1994).

ACKNOWLEDGEMENTS. We thank L. Moresi for providing the CITCOM software on which our fault algorithm is built, and U. Christensen, C. Gable and M. Simons for helpful reviews. This work was supported by the US NSF and the David and Lucile Packard Foundation.

CORRESPONDENCE should be addressed to S.Z. (e-mail: zhong@gps.caltech.edu).

A genome-wide search for quantitative trait loci underlying asthma

Susan E. Daniels*, Sumit Bhattacharya*, Alan James†, Nicholas I. Leaves*, Alan Young*‡, Michael R. Hill§, Jennie A. Faux§, Gerrard F. Ryan†, Peter N. le Souef||, G. Mark Lathrop*, A. William Musk‡ & William O. C. M. Cookson*§

* Wellcome Trust Centre for Human Genetic Disease, University of Oxford, Windmill Road, Oxford OX3 7BN, UK

† Department of Respiratory Medicine, Sir Charles Gairdner Hospital, Perth, Western Australia, Australia

‡ CTSU, Harkness Building, Radcliffe Infirmary, Oxford OX2 6HE, UK

§ Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

|| University Department of Paediatrics, Princess Margaret Hospital, Perth, Western Australia, Australia

ASTHMA now affects one child in seven in the United Kingdom¹. Most cases (95%) of childhood asthma are associated with atopy, the immunoglobulin E (IgE)-mediated familial syndrome of allergic asthma, eczema and rhinitis. Segregation analysis has consistently suggested the presence of major genes influencing atopy and IgE levels^{2–4}, with the expectation that these genes may be identified by positional cloning or the examination of candidate genes. Here we report the results of a genome-wide search for linkage to one qualitative and four quantitative traits associated with allergic (atopic) asthma. We have identified six potential linkages ($P < 0.001$), five of which are to quantitative traits. Monte Carlo simulations show that 1.6 false-positive linkages at this level of significance would be expected from the data. One linkage, to chromosome 11q13, has been established previously⁵. Three of the new loci show evidence of linkage to a second panel of families, in which maternal effects and pleiotropy of linked phenotypes are seen. The results demonstrate the extent and the complexity of the genetic predisposition to asthma.

Asthma is usually recognized epidemiologically by standard symptom questionnaires or by physician diagnosis⁶. The non-specific bronchial hyper-responsiveness that accompanies

asthma is measured by the dose response of airflow to a bronchoconstrictor such as histamine or methacholine. Atopy is detected by skin prick tests, by measurement of specific serum IgE titres against allergens using radioallergosorbent test (RAST) or enzyme-linked immunosorbent assay (ELISA) techniques, or by quantifying the total serum IgE. Atopic disease is accompanied by blood eosinophilia, and eosinophils are prominent in asthmatic airways.

The prevalence of atopy in Western populations is between 40 and 50%⁷⁻⁹. The recruitment of families more than one asthmatic member results in samples in which 70% or more subjects are atopic, making it much more difficult to detect linkage¹⁰. For this reason, the 80 families in the genome screen (from Busselton in Western Australia) were subselected from a population sample to include non-atopic members. We also reasoned that detection of linkage to quantitative traits would be enhanced by including subjects in normal and abnormal ranges of the trait distributions.

We looked for linkage to four quantitative parameters: the total serum IgE, the skin test index (STI), the peripheral blood eosinophil count, and bronchial responsiveness to methacholine (slope)¹¹. A RAST index was also calculated, but gave no additional information to the STI and so was not included in the analysis described below.

Differing indices of atopy may be increased in the same family^{12,13}, and RAST and skin-test responses reach a peak later in childhood than the total serum IgE, and decline at a slower rate thereafter⁹. To account for this heterogeneity of phenotype (pleiotropy), the categorical trait of 'atopy' was used, based on a combination of STI, RAST index, and total serum IgE. Linkage to quantitative traits was tested by the Haseman-Elston sib-pair technique¹⁴, and to atopy by affected sib-pair methods.

The 80 families surveyed contained a total of 203 offspring forming 172 sib-pairs. The mean age of the children was 12.6 ± 1.3 (s.e.) years, their geometric mean IgE was 55.7 ± 1.1 kU⁻¹, and their mean STI was 4.0 ± 0.41 mm; 12% of the children were asthmatic, and 25% admitted to wheeze; 52% were atopic, 36% were non-atopic, and 12% had an intermediate phenotype as defined; the latter were classified as unknown. The sib-pairs subdivided into 44 pairs in which both siblings were atopic, 87 pairs in which one sibling was affected, and 41 in which both were unaffected or of intermediate phenotype.

The families were initially screened with 269 markers (253 autosomal and 16 X-linked). A genome map was constructed from these data, and the order and distances between the markers were compatible with previously published values^{15,16}. Based on these families, the markers had an average heterozygosity of 75%, and spanned approximately 3,100 centimorgans (cM). After initial statistical analysis, none of the X chromosome markers showed evidence of linkage to the five phenotype traits, and the remainder of this analysis therefore deals exclusively with autosomal markers.

Six regions of potential linkage ($P < 0.001$) to autosomal markers were detected with one or more phenotypes on chromo-

TABLE 1 Positive results of genome search (empirical P values)

Marker	θ	Log _e IgE	Skin test index	Log _e eosinophil count	Log _e slope	Atopy
D4S171	0.01				$P < 0.05$	
D4S1540	0.10				$P < 0.05$	
D4S426					$P < 0.0005$	
D6S260	0.21			$P < 0.05$		
D6S276	0.03	$P < 0.05$		$P < 0.0001$		
TNFA	0.03					
D6S273	0.06			$P < 0.05$		
D6S291	0.10					$P < 0.005$
D6S271						$P < 0.05$
D7S484	0.00			$P < 0.05$	$P < 0.0005$	
D7S2250	0.03	$P < 0.005$		$P < 0.05$	$P < 0.05$	
D7S528					$P < 0.05$	
FCERB	0.10	$P < 0.0005$	$P < 0.00005$			
FGF3	0.05					
D11S96	0.09	$P < 0.005$				
D11S901		$P < 0.01$				
D13S262	0.04					$P < 0.05$
D13S270	0.04					$P < 0.005$
D13S153	0.20					$P < 0.001$
D13S170						$P < 0.05$
D16S421	0.11	$P < 0.05$				
D16S515	0.12	$P < 0.05$			$P < 0.05$	
D16S516	0.00	$P < 0.05$				
D16S289	0.06	$P < 0.0005$			$P < 0.05$	
D16S507	0.03	$P < 0.005$			$P < 0.05$	
D16S505		$P < 0.05$				

Only regions showing linkage to one or more phenotypes with $P < 0.001$ are shown. The recombination fraction (θ) is to the subsequent marker.

somes 4, 6, 7, 11, 13 and 16, respectively (Table 1). For the regions on chromosomes 4, 7 and 16, the test statistic met the more stringent criterion of $P < 0.0005$, and for two regions, on chromosomes 6 and 11, P was < 0.0001 . Removal of sib-pair differences that were greater than 3 standard deviations from the mean did not significantly alter the P values for these loci.

The preliminary identification of regions of potential linkage used correlated phenotypes with markers spanning most of the genome. Because of multiple testing, the true significance values will be greater than the normative values given in Table 1. To judge the global significance of the results, a Monte Carlo procedure was applied to calculate the number of regions that would be expected to reach these levels of significance in the absence of any linkages (that is, the number of false-positive linkages). The procedure has advantages over traditional approaches to correcting nominal significance levels because it allows for the informativeness and distance between markers that were actually characterized in the genome-wide search. The expected number of regions of false-positive linkage was 1.6 for $P < 0.001$ (compared with 6 observed linkages), 0.8 for $P < 0.0005$ (5 observed linkages) and 0.2 for $P < 0.0001$ (2 observed linkages), so the observed number of regions meeting these statistical criteria for linkage was much larger than would be expected by chance. In 35,000 simulations, $P < 0.001$ was observed in less than 2% of replicates, indicating that the probability of seeing six or more such regions without a true linkage was < 0.005 .

Additional markers were studied on the six chromosomes that showed evidence for linkage (Table 1). Although the evidence in favour of linkage was not increased, in all instances at least two of the markers from each region had significance of $P < 0.05$.

Chromosomes 11 and 16 exhibited potential linkage to IgE levels. On chromosome 11 the marker FCERB, a microsatellite in

the β -chain of the high-affinity receptor for IgE (Fc ϵ RI- β), also showed linkage to the STI, but this was not observed on chromosome 16. The regions on chromosomes 4 and 7 were linked to bronchial responsiveness, whereas the region on chromosome 6 (near the class I genes of the major histocompatibility complex (MHC)) was linked to eosinophil counts. Weaker evidence of linkage ($P < 0.01$) was found with other phenotypes for the markers on chromosomes 6 and 7 (Table 1).

Markers from chromosome 13, around D13S153, showed evidence of linkage to the atopy phenotype in affected sib-pair analyses (Table 1). The sharing of maternal and paternal alleles identified by descent (IBD) was 46, 1 and 20, 0.

The markers showing $P < 0.001$ for linkage were tested for replication in an additional group (panel B) of 77 nuclear and extended families, recruited from clinics in the United Kingdom, who had previously been used to map atopy on chromosome 11q13⁵. These families had been characterized by the same protocols as the Busseton families, except that bronchial responsiveness and eosinophil counts had not been measured. They contained 215 offspring (268 sib-pairs), of which 61% were atopic and 56% asthmatic, the high values reflecting their origin through clinics. There were 121 atopic-affected sib-pairs, and 70 asthmatic sib-pairs.

Linkage of asthma to FCERB ($P = 0.003$) and to D16S289 ($P = 0.03$) was seen in the panel B families (Table 2). Linkage with atopy was found to D13S153 ($P = 0.003$). The sib-pair sharing of D13S153 alleles in these subjects was 65 IBD = 1, 36 IBD = 0 ($P = 0.003$). The sharing for the combined data was 111 IBD = 1, 56 IBD = 0. The combined Hodge-weighted statistic was 4.158 ($P = 0.000016$). The empirical P value for linkage to D13S153, calculated after 500,000 estimations, was < 0.00001 .

Many studies have shown that maternal atopy and asthma are more likely to be transmitted to children than is paternal atopy¹⁷⁻²⁰. We have not tested for maternal linkage in the genome screen families because of the difficulties induced by a further set of multiple comparisons. However, maternal linkage of atopy to chromosome 11q markers has previously been demonstrated in the panel B families^{5,10}, and so potentially linked markers identified in the genome screen were tested for parent of origin effects (Table 2).

D4S426, FCERB and D16S289 showed significant differences in linkage between maternal and paternal alleles (Table 2). Linkage to maternal meioses was seen between D4S426 and atopy ($P < 0.05$) and the total serum IgE ($P < 0.001$), and between D16S289 and atopy ($P < 0.01$) and asthma ($P < 0.001$). As already observed, FCERB showed strong maternal linkage to atopy^{5,10}, but a strong linkage to asthma was also seen ($P < 0.00001$) which has not been previously noted. The finding of maternal effects at several loci favours immunological interactions between mother and child, rather than genetic imprinting or anticipation, as a cause of these phenomena.

It is not clear if linkage of different traits to different regions of the genome indicates that particular genes affect specific traits. However, a previously established region of genetic linkage, to chromosome 11q13, may be used as an exemplar^{13,21-24}. In the genome screen, the 11q13 marker FCERB was linked to the IgE and the STI. In other studies, FCERB or neighbouring markers have been linked to the composite phenotype of atopy^{13,21,22} and to bronchial responsiveness apparently independently of atopy²⁴. Similarly, D4S426 and D16S289 showed linkage to different phenotypes in genome screen and the panel B families. Although D13S153 showed linkage to the same phenotype of atopy in both Australian and British families, linkage between the total serum IgE and a polymorphism in the esterase D protein, near D13S153, has previously been suggested, with a lod score of 2.07 (ref. 25).

The pleiotropy of phenotypes showing linkage to these loci are consistent with the known variability of atopy and asthma. Pleio-

TABLE 2 Testing for linkage in panel B families: combined and maternal meioses

Locus	IgE	Skin test index	Atopy	Asthma
D4S426	(P < 0.001)*		(P < 0.05)*	
D6S276				
D7S484				
FCERB				P < 0.005
D13S153			(P < 0.0001)**	(P < 0.00001)*
D16S289			P < 0.005	
			(P < 0.01)*	P < 0.05
				(P < 0.001)*

Maternal meioses are given in parentheses. Statistical significance values are as follows: * $P < 0.01$ paternal/maternal differences; ** $P < 0.001$ paternal/maternal differences.

tropy may be attributable to environmental events, such as exposure to allergens, and to the age and selection of subjects. The relative distribution of quantitative traits in normal and abnormal ranges may also have differentially influenced the ability to detect linkage.

The cytokine gene cluster on chromosome 5q31 has previously been reported to be linked to serum IgE in Amish families²⁶, and the result confirmed in Dutch asthmatic families²⁷. In our study we did not see linkage to either the serum IgE or to bronchial responsiveness with D5S393, D5S210 or D5S410 from the cytokine gene cluster, although we are currently typing multiple markers within this 20-cM region.

In summary, both the Monte Carlo simulations and the replication of some positive results in a second set of subjects suggest that we have identified regions of true linkage. Replication of linkage to complex traits is to be expected to be difficult²⁸, and testing for replication in other data sets and with a wider range of markers is still required for the linkages to chromosomes 6 and 7. It is unlikely that all important loci predisposing to atopy and asthma have been mapped, and larger studies and the pooling of data from several centres are required. □

Methods

Subjects. The primary genome screen involved 364 subjects in 80 nuclear families subselected from a population sample of 230 families from Busseton in Western Australia^{4,29}. Clinical testing took place over three months in winter, to control for pollen exposure. Young families were recruited to reduce the effects of age and smoking on the IgE levels and lung function. Families in the genome screen included both atopic and non-atopic members, and sibships of three or greater were not exclusively atopic or non-atopic.

Phenotypes. Skin tests to house-dust mite (HDM) and mixed grass pollen (minus the response of negative controls), specific IgE titres to HDM and timothy grass, and the total serum IgE were measured²⁹. A 'skin test index' (STI) was calculated as the sum of the skin-prick test results to HDM and grass mix; 95% of individuals in this population who were atopic reacted to either HDM or grass pollen, or both. Bronchial responsiveness to (up to 12 μ mol) methacholine was measured²⁹. The slope of the dose-response curve was calculated: (pre-dose forced expiratory volume in one second (FEV1) - last FEV1)/cumulative dose of methacholine. A constant of 0.01 was added to each measurement, to allow log₁₀ transformation when the slope was ≤ 0 . Eosinophils in peripheral blood were Coulter-counted and the values log₁₀-transformed before analysis. Atopy was defined as STI > 5 mm, or a RAST score to HDM and timothy grass > 2 , or a total serum IgE $>$ the 7th decile of the age-corrected population. 'Normal' was defined as an STI of 0 and a RAST index of 0, and a total serum IgE $<$ the 7th decile of the age- and sex-matched population. Intermediate phenotypes were classified as unknown. The subjects were administered a modified British MRC questionnaire²⁹. Asthma was defined as a positive answer to the questions "Have you ever had an attack of asthma?" and "If yes, has this happened on more than one occasion?"

Genotypes. Genomic DNA was isolated from peripheral blood leucocytes by phenol/chloroform extraction. We typed 22 dinucleotide repeat markers by radioactive techniques²⁵ (marker names italicized in Table 1). For 274 markers, 50 ng of each DNA sample was genotyped by fluorescence-based semi-automatic methods¹⁵. Allele sizes were determined using the GENOTYPER software¹⁵. Data were converted into numbered alleles using the GAS (version 2.0) program (A. Young, unpublished). Allele size differences within families rather

than exact alleles were determined. Non-mendelian marker data were flagged by the programs GAS and UNKNOWN. Potentially incorrect genotypes were re-examined and retested if necessary. Two-point lod scores between markers and marker order were compared with published maps^{15,16}.

Statistical analysis. Sib-pairs were included in the linkage analysis only when marker genotypes were known for both parents. Linkage was sought to atopy by affected sib-pair analysis. Sibships containing more than one pair were weighted³⁰. Linkage to quantitative traits was calculated by the method of Haseman and Elston¹⁴, in which the probabilities that sib-pairs have two, one or no alleles identical by descent are regressed on the square of the sib-pair trait differences¹⁴. Linkage was examined further in the region of markers where P was <0.001 . For Table 1, linkage was presented with $P < 0.01$ or $P < 0.05$ if a marker is <15 cM from a marker with a $P < 0.001$. The P values for linkage with the quantitative trait loci were assessed from the regression analysis of variance (ANOVA), which assumes independence of the sib-pairs. In regions of potential linkage ($P < 0.001$), the P values were validated by empirical calculations in which this assumption is not required. The empirical statistics shown in Table 1 were obtained by simulating the segregation of markers to offspring conditional on the observed parental genotypes without linkage to any of the traits. The empirical P values are the proportion of instances in which the linkage statistic equalled or exceeded its observed value in 500,000 simulated replicates. The frequency of false-positive linkages was assessed by a Monte Carlo simulation, in which replicate sets of observations were simulated under the absence of linkage with mendelian segregation of the offspring genotypes conditional on the observed parental genotypes. Linkage between markers was incorporated with recombination distances equal to those estimated from the data map obtained from this study, with offspring simulated under the assumption of lack of interference. The linkage statistics were calculated for each marker locus with the offspring phenotype values as observed. The number of regions in which the linkage statistic for one or more phenotypes exceeded the assigned critical value was calculated over 35,000 replicates. Markers were assumed to detect different regions of linkage if all of the P values were non-significant over a region of >15 cM between two markers with significant linkage.

Replication. Panel B families were recruited from clinics in the United Kingdom⁵. Each family contained at least one asthmatic identified as a proband. Phenotyping was as for the Busselton families, except that eosinophil counts had not been performed and bronchial responsiveness had not been measured. The families were tested with markers that showed potential linkage ($P < 0.001$) to any quantitative or qualitative trait in the Busselton families. Genotyping and linkage analysis were as for the Busselton families. In addition, loci were tested for the presence of significant differences in linkage between maternally and paternally derived alleles; if a significant difference was found, the significance of linkage to maternal alleles was estimated.

Received 4 June; accepted 13 August 1996.

- Strachan, D. P., Anderson, H. R., Limb, E. S., O'Neill, A. & Wells, N. *Arch. Dis. Child.* **70**, 174–178 (1994).
- Gerrard, J. W., Rao, D. C. & Morton, N. E. *Am. J. Hum. Genet.* **30**, 46–58 (1978).
- Borecki, I., Rao, D. C., Lalouel, J. M., McGue, L. & Gerrard, J. W. *Genet. Epidemiol.* **2**, 327–338 (1985).
- Dizier, M. H. et al. *Genet. Epidemiol.* **12**, 93–105 (1995).
- Sandford, A. J. et al. *Lancet* **341**, 332–334 (1993).
- O'Connor, G. T. & Weiss, S. T. *Am. J. Respir. Crit. Care Med.* **149**, S21–28 (1994).
- Hofford-Strevens, V. et al. *J. Allergy Clin. Immunol.* **73**, 516–522 (1984).
- Peat, J. K., Britton, W. J., Salome, C. M. & Woolcock, A. J. *Clin. Allergy* **17**, 271–281 (1987).
- Cline, M. G. & Burrows, B. B. *Thorax* **44**, 425–432 (1989).
- Moffatt, M. F. et al. *Clin. Exp. Allergy* **22**, 1046–1051 (1992).
- O'Connor, G., Sparrow, D., Taylor, D., Segal, M. & Weiss, S. *Am. Rev. Respir. Dis.* **136**, 1412–1417 (1987).
- Cookson, W. O. C. M. & Hopkin, J. M. *Lancet* (i), 86–88 (1988).
- Young, R. P. et al. *J. Med. Genet.* **29**, 236–238 (1992).
- Haseman, J. K. & Elston, R. C. *Behav. Genet.* **2**, 3–19 (1972).
- Reed, P. W. et al. *Nature Genet.* **7**, 390–395 (1994).
- Weissenbach, J. et al. *Nature* **359**, 794–801 (1992).
- Magnusson, C. G. *Allergy* **43**, 241–251 (1988).
- Halonon, M. et al. *Am. Rev. Respir. Dis.* **146**, 866–870 (1992).
- Åberg, N. *Clin. Exp. Allergy* **23**, 829–834 (1994).
- Kuehr, J. et al. *Clin. Exp. Allergy* **23**, 600–605 (1993).
- Cookson, W. O. C. M., Sharp, P. A., Faux, J. A. & Hopkin, J. M. *Lancet* (i), 1292–1295 (1989).
- Shirakawa, T. et al. *Clin. Genet.* **46**, 125–129 (1994).
- Collée, J. M. et al. *Lancet* **341**, 936 (1993).
- Herwerden, L. et al. *Lancet* **346**, 1262–1265 (1995).
- Elberg, H., Lind, P., Mohr, J. & Nielsen, L. S. *Cytogenet. Cell. Genet.* **40**, 622 (1985).
- Marsh, D. G. et al. *Science* **264**, 1152–1155 (1994).
- Myers, D. A. et al. *Genomics* **23**, 464–470 (1994).
- Suarez, B. K., Hampe, C. L. & Van Eerdewegh, P. in *Genetic Approaches to Mental Disorders* (eds Gershon, E. S. & Cloninger, C. R.) 23–46 (American Psychiatric, Washington DC, 1994).
- Hill, M. R. et al. *BMJ* **311**, 776–779 (1995).
- Hodge, S. E. *Genet. Epidemiol.* **1**, 109–122 (1984).

ACKNOWLEDGEMENTS. S.E.D. and S.B. contributed equally to this work. We thank the Busselton families and all those who helped to study them, and the Medical Research Council for providing fluorescence-based microsatellite primer sets; non-fluorescent genotyping was in collaboration with INSERM U358. This work was supported by the Wellcome Trust. Additional funding came from the National Asthma Campaign. G.M.L. holds a Wellcome Trust principal fellowship, and W.O.C.M.C. is a Wellcome senior clinical research fellow.

CORRESPONDENCE and requests for materials should be addressed to W.O.C.M.C. (e-mail: wocc@vax.ox.ac.uk).

Positional cloning of a global regulator of anterior-posterior patterning in mice

Armin Schumacher, Cynthia Faust & Terry Magnuson

Department of Genetics, Case Western Reserve University, Cleveland, Ohio 44106, USA

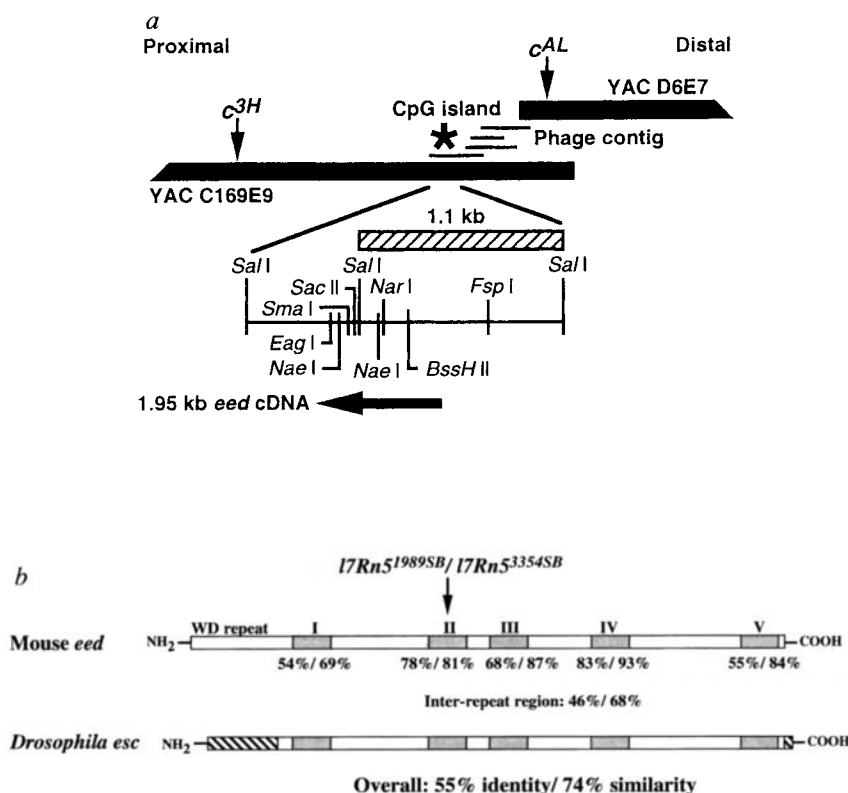
ANTERIOR-POSTERIOR (A-P) patterning is of fundamental importance throughout vertebrate embryonic development. Murine members of the *trithorax* (*trx*) and *Polycomb* group (*Pc-G*) of genes regulate A-P patterning of segmented axial structures^{1–4}, demonstrating conserved upstream regulation of homeotic pathways between *Drosophila* and mouse. Here we report the positional cloning of a classical mouse mutation, *eed* (for *embryonic ectoderm development*), which is the highly conserved homologue of the *Drosophila Pc-G* gene *esc* (for *extra sex combs*), a long-term repressor of homeotic genes⁵. Mutants homozygous for a null allele of *eed* display disrupted A-P patterning of the primitive streak during gastrulation. Mutant embryos lack a node, notochord and somites, and there is no neural induction⁶. In contrast to absence of anterior structures, extra-embryonic mesoderm is abundant. Mice carrying a hypomorphic *eed* mutation exhibit posterior transformations along the axial skeleton. These results indicate that *eed* is required globally for A-P patterning throughout embryogenesis.

The recessive, embryonic-lethal mutation *eed* maps to the albino-deletion complex on chromosome 7 (ref. 7). Previous genetic and molecular analyses positioned *eed* within a 150-kilobase (kb) interval between the *c^{3H}* and the *c^{AL}* deletion breakpoints⁷ (Fig. 1a). Here we have isolated *eed* from a CpG island (that is, a cluster of methylation-sensitive rare-cutter sites) located approximately 40 kb proximal to the *c^{AL}* deletion breakpoint⁷. Molecular access was obtained by chromosomal walking from the proximal end of yeast artificial chromosome (YAC) D6E7 (Fig. 1a). A 1.1-kb *SalI* fragment from the CpG island was used to isolate overlapping clones from a mouse embryonic day 7.5 (E7.5) cDNA library generating a 1.95-kb cDNA (Fig. 1a). The candidate cDNA was validated as the *eed* gene by analysing two non-complementing ethyl-N-nitrosurea (ENU)-induced alleles (*l7Rn5^{3354SB}* and *l7Rn5^{1989SB}*) at the *eed* locus^{8,9}. The *l7Rn5^{3354SB}* and *l7Rn5^{1989SB}* mutations represent null¹⁰ and hypomorphic alleles⁹, respectively. DNA sequence analysis of reverse transcription-polymerase chain reaction (RT-PCR) products from homozygous E8.5 *l7Rn5^{3354SB}* embryos showed a transition of thymine at position 1040 to cytosine (T¹⁰⁴⁰ → C) (Fig. 1a). This base exchange was absent in an ENU-induced allele, *l7Rn5^{4234SB}*, which *trans*-complements the *l7Rn5^{3354SB}* allele⁹. Polymorphism can be excluded because both mutations were generated on the same BALB/cR1 background⁸. As a second co-isogenic control, RT-PCR products from E12.5 homozygous *l7Rn5^{1989SB}* embryos also showed the expected wild-type T¹⁰⁴⁰. The sequence alteration in *l7Rn5^{1989SB}* was identified as a T¹⁰³¹ → A transversion which was absent in both *l7Rn5^{4234SB}* and *l7Rn5^{3354SB}* (Fig. 1a).

Conceptual translation revealed an open reading frame of 441 amino acids, containing five tandemly repeated WD motifs (Fig. 1a). A potential PEST sequence¹¹ is defined by residues 22 to 52 of Eed protein (Fig. 1a). Two discrete clusters of basic amino acids encompassing a nine-residue spacer region (RK-X₉-KKKKX, residues 65 to 80; Fig. 1a) are reminiscent of a bipartite nuclear targeting signal¹². Eed does not display any canonical DNA-binding motif. Three putative Asn-linked glycosylation sites (NXS/T)¹³ at amino-acid positions 157, 283 and 349 (Fig. 1a) are compatible with a potential non-nuclear isoform of Eed generated by loss of the nuclear targeting signal through alternative splicing (A. Schumacher and T. Magnuson, unpublished

FIG. 1 a, Positional cloning of *eed*. A YAC contig (dark bars) was previously constructed from *D7Cwr11D* across the 150-kb *eed* region demarcated by the c^{3H} and the c^{AL} deletion breakpoint⁷ (vertical arrows). Chromosomal walking from the proximal end of YAC D6E7 yielded access to a CpG island characterized by 11 methylation-sensitive rare-cutter sites within a 1.7-kb genomic region. The 1.1-kb *SalI* fragment (hatched bar) was used to isolate overlapping clones corresponding to a 1.95-kb cDNA extending in proximal orientation (arrow). A 441-amino acid open reading frame predicts a protein of relative molecular mass 50,000 (M_r 50K) and a pI of 6.6. Five WD motifs (shaded), a potential PEST sequence (underlined), a potential bipartite nuclear targeting signal (dashed underlined), and three potential Asn-linked glycosylation sites (boxed) are indicated. ENU-induced point mutations in the second WD motif are denoted by arrows. **b**, Comparison of mouse *Eed* and *Drosophila melanogaster* *Esc* (refs 17–19) sequences. Five WD repeats (I–V) in *Eed* and *Esc* are highlighted by stippled boxes. Percentage amino-acid identity/percentage similarity are indicated for the overall sequence comparison, along with the individual WD motifs and the inter-repeat region. Similarity rules for protein sequence alignments are derived from positive scoring pairs in the BLOSUM62 substitution matrix. Calculation of sequence conservation excluded divergent N- and C-terminal residues (hatched boxes). Arrow indicates position of ENU-induced point mutations.

METHODS. The proximal end of YAC D6E7, previously isolated by inverse PCR²⁰, was used to screen a 129/Sv lambda phage genomic library (Stratagene). The 1.1-kb *SalI* fragment was used to screen a mouse oligo dT-primed E7.5 cDNA library in λ ZAP II (provided by J. D. Gearhart). Phagemids were isolated from plaque-purified clones by *in vivo* excision. Unidirectional deletions were constructed using exonuclease III and mung bean nuclease, and nested cDNA inserts were sequenced using vector primers. Total RNA was isolated from E8.5 and E12.5 homozygous mutant



results). The $T^{1040} \rightarrow C$ transition in $17Rn5^{3354SB}$ changed a Leu to a Pro residue at position 196 (L196P) in the second WD motif (Fig. 1a). Although this is a conservative substitution with regard to isoelectrical properties, Pro is unique among amino acids in causing sharp kinks in chain conformation¹⁴. Because WD motif-mediated protein–protein interaction is likely to be conformation sensitive¹⁵, the constraints imposed by the Pro substitution in the second WD motif provided a convincing molecular basis for the null phenotype. The $T^{1031} \rightarrow A$ transversion at position 1031 in the hypomorphic $17Rn5^{1989SB}$ allele substituted a polar Asn residue for the hydrophobic Ile at position 193 in the second WD motif (Fig. 1a). Asn imposes much lower steric restrictions on the polypeptide backbone than does Pro¹⁶, and so is compatible with the relative phenotypic refractivity.

Database searches revealed highly significant protein sequence similarity to *extra sex combs* (*esc*)^{17–19}, a *Drosophila* *Pc-G* gene required for long-term transcriptional repression of homeotic genes⁵. Sequence alignment of *Eed* and *Esc* proteins revealed 55% identity and 74% similarity (Fig. 1b). Conservation between the WD motifs (46% identity, 68% similarity) is almost as high as that within the WD motifs (67% identity, 82% similarity). This high evolutionary conservation spans 83% of the *Eed* sequence, and is not interrupted by a single gap or insertion. The amino-terminal region has diverged considerably between *Eed* and *Esc*.

Developmental RT–PCR showed a large maternal messenger RNA pool in oocytes; preimplantation embryos contained only

$17Rn5^{3354SB}$ and $17Rn5^{1989SB}$ embryos, respectively. Homozygous mutant PO $17Rn5^{4234SB}$ newborn RNA served as a co-isogenic control. First-strand reverse transcription was done with random hexanucleotide priming and M-MLV reverse transcriptase. cDNAs were PCR-amplified using primers flanking the open reading frame (positions 429–453 and 1830–1852), subcloned and sequenced using internal primers. The $17Rn5^{3354SB}$ mutation disrupted an *AluI* site that cosegregated with the mutant phenotype. The $17Rn5^{1989SB}$ mutation was verified by sequencing of RT–PCR clones from 10 embryos.

minute amounts of *eed* transcripts, and these were probably remnants of the maternal *eed* pool (Fig. 2a). Embryonic *eed* expression was found to begin at E5.5, followed by high expression throughout development (Fig. 2a), and northern blot analysis indicated constitutive expression of *eed* in adult tissues (Fig. 2b). Two highly expressed transcripts of approximately 1.7 kb and 2.2 kb were detected in all lanes. The relative intensity of the two transcripts was dependent on tissue type. The 2.2-kb band is consistent with a full-length cDNA clone when a poly(A) tail of 100–200 base pairs is taken into account. The smaller band probably derives from complex alternative splicing (A. Schumacher and T. Magnuson, unpublished results). mRNA *in situ* hybridization analysis of sectioned E6.5 to E8.5 wild-type embryos confirmed that *eed* is expressed ubiquitously (data not shown).

It has been shown that *esc* and other *Pc-G* genes, in conjunction with transcriptional activators of the *trx* group, are responsible for the long-term maintenance of spatially restricted expression of homeotic genes along the A–P body axis⁵. The current model favours transcriptional regulation of homeotic gene expression by multimeric protein complexes modifying higher-order chromatin structure²⁰. *Drosophila* embryos deficient for both maternal and embryonic *esc* display transformation of all body segments to an eighth abdominal segment identity and derepression of homeotic genes of the *Bithorax* complex^{21,22}. Thus, on the basis of sequence homology to *Drosophila esc*, mutations in *eed* should cause spatial derepression of *Hox* genes and homeotic transformations of axial

skeletal structures. Indeed, analysis of 179 newborn skeletons from *inter se* crosses of the hypomorphic *17Rn5^{1989SB}* allele indicated conserved function between *eed* and *esc* by showing a dose-dependent increase in posterior transformations along the entire A–P axis (Fig. 3a–f, Table 1).

Phenotypic analysis of the *eed* null allele⁶ demonstrates an additional role of *eed* earlier in development than that predicted from *esc* function. A–P patterning of genetically determined segmental structures is maintained by *esc*, but *esc* does not regulate axial polarity in the unsegmented egg²¹. In contrast, disruption of the unsegmented primitive streak during gastrulation⁶ reflects *eed* involvement in regulating A–P patterning before segmentation, which is first observed during somitogenesis in the mouse. Unlike the regulation of segmental identity, A–P patterning of the primitive streak does not involve homeotic gene expression from the *Hox* clusters, as these genes are not expressed before the late-streak stage²³, a developmental stage reached by very few *eed* null embryos⁶. Indeed, the degree of *Otx2* anterior localization in the most advanced *eed* mutant embryos demonstrates development only to the mid-streak stage (Fig. 2c–f). Additional evidence for lack of ectoderm development comes from the absence of neural markers such as *Krox20* and *Engrailed* (data not shown). Paraxial mesoderm formation also did not occur, as indicated by the absence of *Mox1* expression (data not shown).

Expression analyses of *eed* null embryos showed that the homeo-domain-containing putative transcription factor *Evx1* is downstream of *eed*⁶. Mid-streak-stage wild-type embryos show a graded

TABLE 1 Penetrance of skeletal transformations in hypomorphic *17Rn5^{1989SB}* mutants

	+/+ (%; n = 86)	<i>17Rn5^{1989SB}/+</i> (%; n = 50)	<i>17Rn5^{1989SB}/17Rn5^{1989SB}</i> (%; n = 43)
Fusion of atlas and dens axis	15	26	93
Anterior tuberculi shifted to C5	9	2	53
Ribs on C7	7	8	26
Six vertebrosteral ribs	0	0	26
Loss of ribs on T13	0	0	14
Fusion of L6 to ilial bones	24	56	100

Homozygous skeletons displayed a high penetrance of transformations with complete penetrance at the posterior end. *eed* dosage sensitivity was apparent in heterozygous skeletons with penetrance of transformations falling between wild-type and homozygous mutant skeletons. Both unilateral and bilateral posterior transformations were included in axial skeletal analysis.

pattern of *Evx1* expression in the primitive streak, with higher expression levels found more posteriorly²⁴ (see also Fig. 2g). Although initial expression of *Evx1* in the early primitive streak of *eed* mutant embryos was normal⁶, the graded posterior to anterior expression of *Evx1* in the primitive streak was never seen (Fig. 2h). Absence of the *Evx1* gradient and high expression throughout the primitive streak could alter the fate of ingressing cells forming excess extra-embryonic mesoderm at the expense of embryonic mesoderm. This hypothesis is based on cell-fate analysis in wild-type embryos showing ingression of prospective extra-embryonic mesoderm cells through the posterior part of the primitive streak²⁵, which displays higher levels of *Evx1* expression²⁴.

FIG. 2 Expression analysis of *eed*. a, Developmental RT–PCR of wild-type embryonic stages. b, Northern blot analysis of mouse adult tissues. c–f, In wild-type embryos, *Otx2* is expressed throughout the epiblast before gastrulation (c), undergoes anterior localization from the mid- (d) to late-streak stage (e), and is eventually confined to the anterior neural ectoderm³⁰. In most E8.5 homozygous *17Rn5^{3354SB}* embryos, *Otx2* transcripts were found throughout the embryonic ectoderm (f, right). Very few mutant embryos displayed some anterior localization (f, left). g, In E7.5 wild-type embryos, *Evx1* is expressed in a graded pattern in the primitive streak with higher expression levels found more posteriorly (arrowhead). h, E8.5 mutant embryos display absence of graded *Evx1* expression in the primitive streak (arrowhead) and an ectopic *Evx1* expression domain in the most proximo-anterior part of the epiblast (arrow). c–h, Anterior is to the left, posterior to the right. Scale bar, 100 µm. METHODS. Pools of 54 oocytes, 18 two-cell, 11 four-cell, 15 eight-cell, five morulae, 13 early blastocyst, or 5 hatched blastocyst-stage embryos were digested with proteinase K. RT–PCR was performed using whole lysates from preimplantation and E5.5 to E8.5 embryos, as well as 1 µg of total RNA from E9.5 to E18.5 embryos. PCR primers correspond to positions 798–818 and 1147–1166, respectively, of the *eed* cDNA, as well as standard *GAPDH* primers. A full-length *eed* cDNA probe was hybridized to a multiple tissue northern (Clontech). Whole-mount mRNA *in situ* hybridization of E6.5 to E7.5 wild-type and E8.5 *eed* embryos was performed using an *Otx2* probe³⁰. A 588-bp cDNA probe corresponding to positions 393–981 of *Evx1* was used for mRNA *in situ* hybridization.

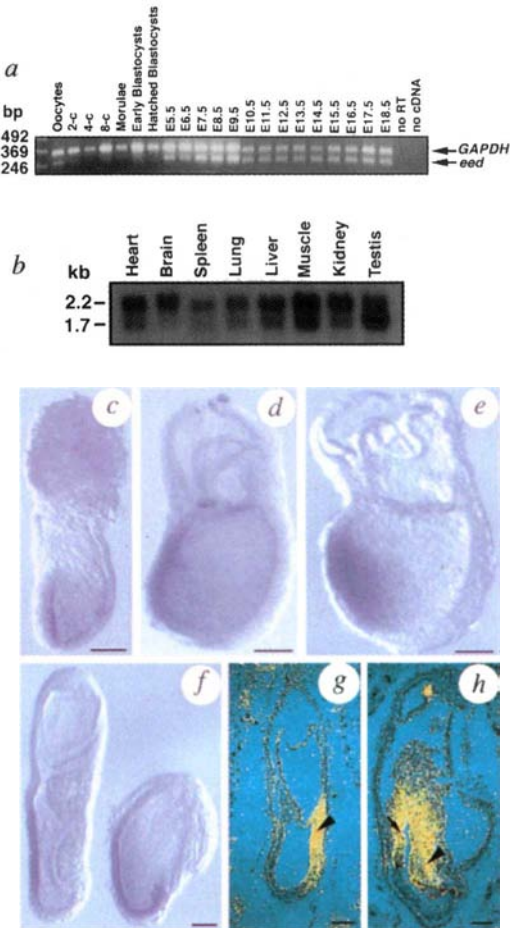
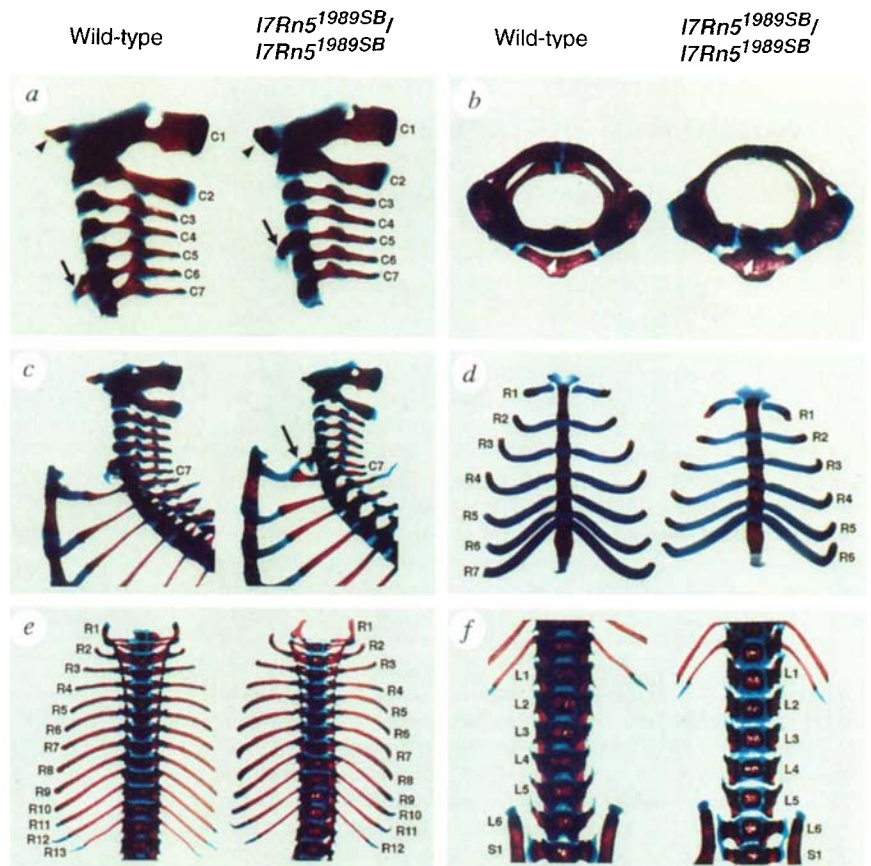


FIG. 3 Posterior skeletal transformations in hypomorphic *17Rn5^{1989SB}* mutants. Wild-type skeletons are shown on the left, and homozygous *17Rn5^{1989SB}* skeletons on the right. a, Lateral view of dissected cervical vertebrae (C1–C7). The anterior arch of the atlas (C1) (arrowhead) is pointed and does not fuse with the axis (C2) in a wild-type skeleton. In a *17Rn5^{1989SB}* skeleton, the atlas is broadened, extends caudally, and fuses with the vertebral body of the axis. The anterior tuberculi, which are located at C6 in the wild type, are shifted anteriorly to C5 in a mutant skeleton (arrow). Thus C5 and C6 acquired a posterior identity similar to C6 and C7, respectively. b, Cranial view into the dissected atlas and axis showing a broad dorsal fusion between the anterior arch of the atlas and the dens axis in a *17Rn5^{1989SB}* skeleton (arrow). This represents a posterior transformation of the atlas, probably deriving from an incomplete regression of the body of the atlas during embryogenesis. c, Lateral view of dissected cervical and upper thoracic region. An extra rib emanating from C7 fuses to the first rib in a *17Rn5^{1989SB}* skeleton, transforming C7 into a thoracic identity (arrow). d, Ventral view of dissected rib cage. Seven vertebrosteral ribs in a wild-type skeleton are denoted by R1–R7. A *17Rn5^{1989SB}* skeleton exhibits only six vertebrosteral ribs, reflecting a more posterior identity of the seventh thoracic vertebra. e, Dorsal view of the thoracic vertebral column with dissected ribs. A wild-type skeleton shows 13 pairs of ribs (R1–R13). In contrast, a mutant skeleton lacks the 13th pair of ribs, representing a lumbar identity of the last thoracic vertebra. f, A homozygous *17Rn5^{1989SB}* skeleton shows bilateral fusion of the sixth lumbar vertebra to the iliac bones, which is a sacral identity. L1–L6 and S1 denote the six lumbar vertebrae and the first sacral vertebra, respectively.



METHODS. PO–P6 whole-mount skeletons deriving from *17Rn5^{1989SB}/c^{ch}* *inter se* crosses were stained as described previously¹. To exclude meiotic recombinations between *eed* and the albino locus (c), D7MIT352 was amplified by PCR from tail DNA samples. This microsatellite marker maps

distal to *eed* between the *c^{110SD}* and *c^{2YPSj}* deletion breakpoints⁷ (A. Schumacher and T. Magnuson, unpublished results), and is polymorphic between the chinchilla and the BALB/cR1 chromosome on which *17Rn5^{1989SB}* was induced⁸.

In contrast, the anterior part of the primitive streak, which gives rise to embryonic mesoderm, contains lower levels of *Evx1* expression²⁴. A potential role for *Evx1* in posterior mesoderm patterning is evident from studies in *Xenopus*. *Xhox3*, the frog homologue of *Evx1*, is similarly expressed in a graded fashion along the A–P axis, and reduction of *Xhox3* function disrupts development of posterior structures²⁶. In contrast, ectopic expression of *Xhox3* in anterior domains inhibits anterior mesoderm patterning²⁷. Similar results have been obtained by ectopic expression of zebrafish *eve1* (ref. 28), and by mutational analysis of the *Caenorhabditis elegans* homologue *vab-7* (ref. 29). Thus it is conceivable that *eed* may be required to establish and/or maintain

an *Evx1* expression gradient, which in turn may implement positional information on mesodermal cells leaving the primitive streak.

We have shown that *eed* regulates A–P patterning throughout the length of the unsegmented primitive streak. Other mouse *Pc-G* genes such as *bmi1* and *mel18* are not essential for A–P patterning of the primitive streak, as mice carrying null alleles of either gene survive to birth^{1,3}. Similarly, embryos deficient for a transcriptional activator of homeotic genes, *Mll*, the mouse homologue of *Drosophila* *trx*, do not demonstrate a disruption of the primitive streak⁴. These data suggest that other, as yet unidentified, transcriptional activator(s) antagonize *eed* function during gastrulation. □

Received 17 April; accepted 22 July 1996.

- van der Lugt, N. M. et al. *Genes Dev.* **8**, 757–769 (1994).
- Alkema, M. J., van der Lugt, N. M., Bobeldijk, R. C., Bems, A. & van Lohuizen, M. *Nature* **374**, 724–727 (1995).
- Akasaka, T. et al. *Development* **122**, 1513–1522 (1996).
- Yu, B. D., Hess, J. L., Horning, S. E., Brown, G. A. & Korsmeyer, S. J. *Nature* **378**, 505–508 (1995).
- Simon, J. *Curr. Opin. Cell Biol.* **7**, 376–385 (1995).
- Faust, C., Schumacher, A., Holdener, B. & Magnuson, T. *Development* **121**, 273–285 (1995).
- Holdener, B. C. et al. *Genomics* **27**, 447–456 (1995).
- Rinchik, E. M., Carpenter, D. A. & Selby, P. B. *Proc. Natl Acad. Sci. USA* **87**, 896–900 (1990).
- Rinchik, E. M. & Carpenter, D. A. *Mamm. Genome* **4**, 349–353 (1993).
- Holdener, B. C., Rinchik, E. M. & Magnuson, T. *Mamm. Genome* **6**, 474–475 (1995).
- Rogers, S., Wells, R. & Reichsteiner, M. *Science* **234**, 364–368 (1986).
- Dingwall, C. & Laskey, R. A. *Trends Biochem. Sci.* **16**, 478–481 (1991).
- Kornfeld, R. & Kornfeld, S. *Annu. Rev. Biochem.* **56**, 615–664 (1987).
- MacArthur, M. W. & Thornton, J. M. *J. Mol. Biol.* **218**, 397–412 (1991).
- Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. *Nature* **371**, 297–300 (1994).
- Stites, W. E. & Pranata, J. *Proteins* **22**, 132–140 (1995).
- Sathe, S. S. & Harte, P. J. *Mech. Dev.* **52**, 77–87 (1995).

- Gutjahr, T. et al. *EMBO J.* **14**, 4296–4306 (1995).
- Simon, J., Bornemann, D., Lunde, K. & Schwartz, C. *Mech. Dev.* **53**, 197–208 (1995).
- Orlando, V. & Paro, R. *Curr. Opin. Genet. Dev.* **5**, 174–179 (1995).
- Struhl, G. *Nature* **293**, 36–41 (1981).
- Struhl, G. & Akam, M. *EMBO J.* **4**, 3259–3264 (1985).
- Deschamps, J. & Wijgerde, M. *Dev. Biol.* **156**, 473–480 (1993).
- Dush, M. K. & Martin, G. R. *Dev. Biol.* **151**, 273–287 (1992).
- Lawson, K. A., Meneses, J. J. & Pedersen, R. A. *Development* **113**, 891–911 (1991).
- Ruiz i Altaba, A., Choi, T. & Melton, D. A. *Dev. Growth Differ.* **33**, 651–669 (1991).
- Ruiz i Altaba, A. & Melton, D. A. *Cell* **57**, 317–326 (1989).
- Barro, O. et al. *Dev. Genet.* **17**, 117–128 (1995).
- Ahringer, J. *Genes Dev.* **10**, 1120–1130 (1996).
- Ang, S. L., Conlon, R. A., Jin, O. & Rossant, J. *Development* **120**, 2979–2989 (1994).

ACKNOWLEDGEMENTS. We thank R. Conlon, P. Harte and members of our laboratory for reading the manuscript and discussions, and U. Rutishauser, H. Hu and K. Gustashaw for help with imaging. A.S. is a recipient of a postdoctoral Feodor Lynen fellowship from the Alexander von Humboldt Foundation. C.F. is a recipient of a postdoctoral fellowship from the American Cancer Society. This work was supported by grants from the NIH and March of Dimes to T.M.

CORRESPONDENCE and requests for materials should be addressed to T.M. (e-mail: trm4@po.cwru.edu).

Functional anatomy of a common semantic system for words and pictures

R. Vandenberghe, C. Price, R. Wise, O. Josephs
& R. S. J. Frackowiak

Wellcome Department of Cognitive Neurology, Institute of Neurology,
12 Queen Square, London WC1N 3BG, UK

THE relationship between the semantic processing of words and of pictures is a matter of debate among cognitive scientists^{1,2}. We studied the functional anatomy of such processing by using positron-emission tomography (PET). We contrasted activity during two semantic tasks (probing knowledge of associations between concepts, and knowledge of the visual attributes of these concepts) and a baseline task (discrimination of physical stimulus size), performed either with words or with pictures. Modality-specific activations unrelated to semantic processing occurred in the left inferior parietal lobule for words, and the right middle occipital gyrus for pictures. A semantic network common to both words and pictures extended from the left superior occipital gyrus through the middle and inferior temporal cortex to the inferior frontal gyrus. A picture-specific activation related to semantic tasks occurred in the left posterior inferior temporal sulcus, and word-specific activations related to semantic tasks were localized to the left superior temporal sulcus, left anterior middle temporal gyrus, and left inferior frontal sulcus. Thus semantic tasks activate a distributed semantic processing system shared by both words and pictures, with a few specific areas differentially active for either words or pictures.

Patients with visual agnosia sometimes perform semantic judgements correctly when concepts are visually presented as words, but not when presented as pictures, despite relatively spared perception of the stimulus surface structure³⁻⁶. Other patients do significantly better in semantic tasks with pictures than with the corresponding words, although they can read words aloud⁷. These and related dissociations⁸⁻¹¹, along with picture-word priming and interference studies in normal subjects¹², have led to divergent cognitive models¹. Semantic representation systems accessed by words and pictures may be segregated, and modality-specific semantic deficits may then be due to degradation of one of these systems⁹. In a contrasting view, only the procedure by which higher-level surface descriptions are mapped onto concepts differs between words and pictures, and the dissociations reflect a functional disconnection between modality-specific presemantic processing and an amodal knowledge system^{13,14}. According to a third model, the distinction between higher-level perceptual presemantic processing and modality-specific semantic processing is artificial, as even associative visual agnosics often also have higher-level perceptual deficits³.

In our experiment, subjects performed semantic or non-semantic judgements on triplets of either words or pictures¹⁵. During one task (Fig. 1a, d), the stimuli within each triplet belonged to the same category, and subjects matched the stimuli for meaning (associative semantics). During the second task (Fig. 1b, e), subjects matched stimuli for the real-life size of their referents (visual semantics)¹⁶. During a third task (Fig. 1c, f), they matched stimuli for physical size (baseline). The behavioural data are shown in Table 1.

We began by determining where word processing significantly differed from picture processing, irrespective of the task performed. We compared pictures with words separately for each of the three tasks, including the non-semantic task. Modality-specific presemantic processing was postulated where regional cerebral blood flow (rCBF) differed significantly ($P < 0.01$) between input modalities in each of these contrasts. Because of the orthogonality

TABLE 1 Behavioural data

	Reaction time (ms)		Accuracy (% correct)	
	mean	s.e.	mean	s.e.
Assoc.(W)	2,705	52	82.8	4
Assoc.(P)	2,635	57	79.5	5
Vis.sem.(W)	2,919	47	89.8	2
Vis.sem.(P)	2,698	50	77.9	3
Baseline(W)	2,639	60	71.0	6
Baseline(P)	2,818	50	71.4	6

The reaction times confirm the advantage of pictures to words during real-life size judgements, as well as the similarity in reaction time between pictures and words during semantic distance judgements¹⁶. A two-way repeated-measures ANOVA by items for the reaction times showed a significant main effect of task ($F(2,1331) = 3.4$; $P < 0.05$), no significant effect of materials, and a significant task by materials interaction ($F(2,1331) = 7.4$; $P < 0.001$). A two-way repeated-measures ANOVA by subject for the accuracies showed a significant main effect of task ($F(2,65) = 4.3$; $P < 0.05$), as the accuracies for the active conditions were higher than for the baseline. There was no significant effect of materials, nor a significant interaction. Abbreviations: Assoc., associative semantics; Vis.sem., visual semantics; P, pictures; W, words.

of these three contrasts, the probability of reaching this criterion by chance is approximately 10^{-6} . Picture-specific processing, irrespective of the task performed, occurred in the right middle occipital gyrus ($42, -72, 4$; BA19/37; Fig. 2b). A nearby or identical area ($34, -72, -8$) has been reported to be more active when subjects passively view nonsense objects than when they view visual noise patterns¹⁷. Word-specific processing, irrespective of the task, occurred in left inferior parietal lobule ($-34, -36, 24$; BA40; Fig. 2a).

We then determined where the semantic network overlapped between words and pictures. Overlap was defined as a significant activation during associative or visual semantic judgements compared with baseline when the tasks were performed using words

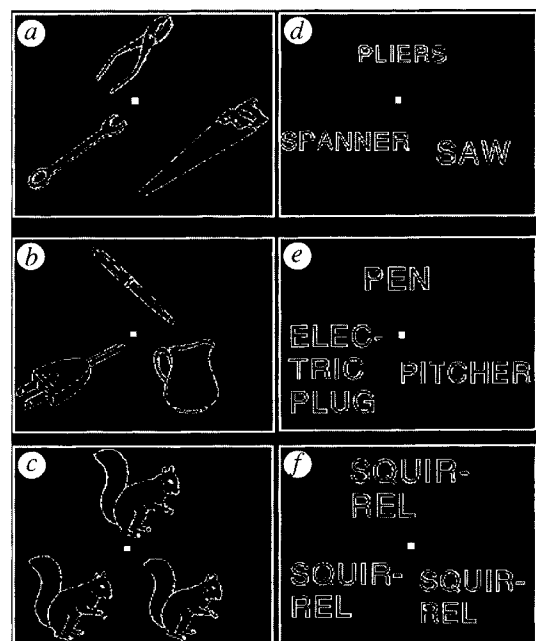


FIG. 1 Examples from stimuli for each of the picture (a-e) and word (d-f) conditions. a, d, Associative semantics; b, e, visual semantics; c, f, baseline conditions.

($P < 0.01$) and when using pictures ($P < 0.01$). Words and pictures shared a distributed semantic system (Fig. 2e and Table 2), consisting of many regions previously described in investigations of word generation and comprehension: the junction between parietal and temporal¹⁸, and between fusiform and inferior temporal cortex¹⁸⁻²⁰, the left middle temporal gyrus^{21,22}, and left inferior frontal gyrus (BA11/47)²³. Whatever the exact semantic operations performed by these areas, they are activated for both input modalities (Table 2). This shared semantic network was activated whichever type of semantic task subjects performed, although left hippocampal and right lateral cerebellar activations were stronger during associative than during visual semantics ($Z = 3.21$ and 3.66 , respectively). These were the only two areas in the entire brain volume searched where the two types of semantic task differed significantly ($P < 0.001$, uncorrected). The common activation pattern for the two types of semantic task may reflect anatomical overlap between an associative semantic and a visual semantic system, or activation of both systems during both semantic tasks.

Finally, we determined where semantic processing differed between words and pictures. We delineated a semantic processing network by comparing all associative semantic tasks to baseline ($P < 0.01$), and then determined within that network the areas where the difference between the semantic tasks and the baseline condition differs significantly between words and pictures ($P < 0.01$). The probability of reaching this conjoint criterion by chance is approximately 10^{-4} . Left superior temporal sulcus ($-58, -30, 0$), left anterior middle temporal gyrus ($-42, -10, -16$) and left inferior frontal sulcus ($-34, 26, 20$) were activated during associative semantics compared with baseline ($Z = 3.69, 3.95$ and 3.30 , respectively), and significantly more when these tasks were performed with words instead of pictures (interaction, $Z = 2.83, 2.49$ and 2.33) (Fig. 2c). Activation during semantic tasks specific to pictures was localized to the left posterior inferior temporal sulcus ($-38, -62, 4$) (main task effect, $Z = 2.37$; interaction, $Z = 2.57$) (Fig. 2d). No modality-specific semantic areas were revealed in the right hemisphere ($P > 0.05$, uncorrected).

None of the modality-specific areas showed any influence of the type of semantic task ($P > 0.05$, uncorrected). The absence of any effect of semantic task content contradicts the hypothesis that these areas would constitute semantic subsystems containing modality-specific knowledge. Instead, these input-specific, task-related increases in rCBF may arise because, compared with visual-perceptual tasks, semantic tasks may lead to enhanced activation of picture-specific structural descriptions²⁴ or word-specific lexical or phonological²⁵ representations. The blood flow increases may also relate to the process of accessing a shared semantic representation from these modality-specific representations^{13,14}.

Our results provide strong evidence for a distributed semantic system shared by both input modalities. In monkeys, inferotemporal cortex and ventral frontal convexity are known to be involved in object recognition^{26,27}. The anatomy and function of the common semantic processing stream we describe suggests that, phylogenetically, when primates acquired language, a pre-existing object-recognition system could have been adapted to attribute meaning to nouns. □

TABLE 2 Common semantic network

		rCBF increase (%)			
		Assoc. Words	Vis.sem. Pictures	Assoc. Pictures	Vis.sem. Pictures
L inf. front. g. (BA45)	-42,22,20	3.9 $Z = 5.36$	3.4	2.4 $Z = 3.45$	1.8
L inf. front. g. (BA11/47)	-16,30,-12	2.7 $Z = 3.34$	1.7	2.0 $Z = 2.33$	0.9
L mid. temp. g. (BA21)	-42,0,-28	2.9 $Z = 3.40$	2.4	1.9 $Z = 2.37$	1.1
L mid. temp. g. (BA21)	-58,-38,-4	3.1 $Z = 4.42$	2.3	1.7 $Z = 2.58$	2.3
L inf. temp. g. (BA20)	-44,-10,-28	3.3 $Z = 4.12$	2.7	1.9 $Z = 2.36$	1.4
L fusiform g. (BA21/37)	-46,-46,-20	2.9 $Z = 3.64$	2.8	2.4 $Z = 3.03$	2.0
L PT (BA19/39)	-40,-70,24	1.8 $Z = 3.20$	1.5	2.1 $Z = 3.46$	1.5
L sup. occ. g. (BA19)	-30,-76,40	3.3 $Z = 4.00$	1.9	2.2 $Z = 2.75$	2.4
L hippocampus (BA34)	-18,-16,-12	2.3 $Z = 4.27$	1.1	1.2 $Z = 2.36$	0.2
vermis	14,-82,-28	3.0 $Z = 4.94$	1.4	2.3 $Z = 3.92$	2.1
R cerebellum	38,-76,-48	4.8 $Z = 3.80$	2.5	4.4 $Z = 3.41$	1.9

In all of these areas, the probability of being activated in the contrast between associative semantics and baseline both for words and for pictures by chance was lower than $P < 10^{-5}$. First row: anatomical name, Talairach coordinates, percentage rCBF difference with respect to baseline, calculated by subtracting the adjusted rCBF in the respective baseline conditions from those in the semantic conditions and dividing by the respective baseline rCBF. Second row: the Z-scores for the subtraction between associative semantics and baseline. Abbreviations: Inf. front. g., inferior frontal gyrus; mid. temp. g., middle temporal gyrus; inf. temp. g., inferior temporal gyrus; PT, parietotemporal junction; sup. occ. g., superior occipital gyrus.

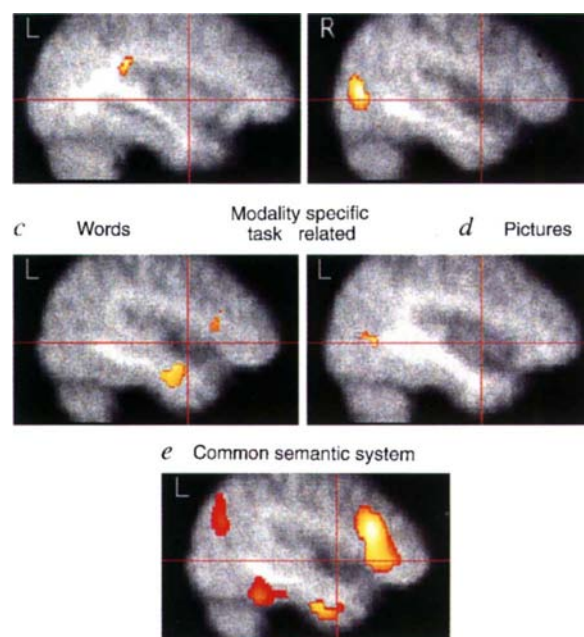


FIG. 2 Sagittal sections through the subjects' normalized averaged brains at $x = -34$ mm (left) (a), at $+42$ mm (right) (b), at -42 mm (left) (c, e), and at -38 mm (left) (d). a, Areas activated when subtracting pictures from words irrespective of task. b, Areas activated when subtracting words from pictures irrespective of task. c, Interaction between materials and task where the difference between associative semantics and baseline is larger for words. d, Interaction between materials and task where the difference is larger for pictures. e, Areas activated when the associative semantic task performed with words is compared with the non-semantic task performed with words and also activated when the same tasks are compared for pictures.

Methods

Subjects. Six right-handed male subjects, aged between 22 and 69 years, participated in this study. They gave their written, informed consent in accordance with the Declaration of Helsinki.

Stimuli. Nearly all pictures belonged to the Snodgrass–Vanderwart stimulus set²⁸. Stimulus triplets were presented on a screen 36.6 deg wide and 29.2 deg high, at 45 cm distance, for 4 s with 500 ms between presentations. Picture size was on average 11 deg, letters were 3.2 deg high and 2.5 deg wide. The sample stimulus was presented at 6.7 deg above the screen centre, the test stimuli at 5.6 (±2) deg below the screen centre and 10.3 (±2) deg to the left and the right.

Tasks. Subjects performed three different matching-to-sample tasks either with pictures (Fig. 1a–c) or with words (Fig. 1d–f). In one task (Fig. 1a, d), they selected the left or right key-press button, according to whether the left or right lower stimulus was closer in meaning to the sample stimulus above. Examples: cow: horse, bear; switch: light bulb, candle; coat: glove, sock; cucumber: tomato, corn. In a second semantic task (Fig. 1b, e), they performed a matching-to-sample task for real-life size¹⁶. Perceptual size cues were minimized. Examples: button: glass, bottle; suitcase: beaver, well; chisel: banana, kite; nail: thimble, salt cellar. The baseline task (Fig. 1c, f) consisted of a matching-to-sample task for physical size on the screen. In this baseline task, the distractor was 6% larger or smaller than the target, and the target was 6% larger or smaller than the sample size. Whether a triplet was presented as either words or pictures was entirely counterbalanced between subjects. Each subject saw each triplet of concepts only once, and each concept was used equally often in each condition.

Data acquisition. The brain was scanned with an ECAT EXACT HR+ PET scanner. The task order was counterbalanced. Each subject was scanned twice in each of 6 conditions. Each run consisted of 19 trials. Images of all 12 conditions were available for each of the 6 subjects.

Data analysis. The scans from each subject were realigned using the first image as a reference. A T1-weighted magnetic resonance image (MRI) was co-registered to the mean PET image for each subject and then stereotactically transformed²⁹ to a standard MRI template in the Talairach and Tournoux space, and the same transformation matrix subsequently applied to the PET images. Data were analysed using a randomized block design with replication and global brain activity as two covariates of no interest³⁰. The search volume went from

$z = -48$ mm to $z = 60$ mm, with a final image resolution (full width of half-maximum) $x = 16.9$ mm, $y = 18.9$ mm, $z = 20.8$ mm.

Received 22 April; accepted 24 July 1996.

1. Caplan, D. *Language: Structure, Processing and Disorders* 63–104 (MIT Press, Cambridge, MA, 1992).
2. Shallice, T. In *The Cognitive Neuropsychology of Language* (eds Coltheart, M., Sartori, G. & Job, R.) 111–127 (Lawrence Erlbaum, Hove, 1987).
3. Farah, M. *Visual Agnosia: Disorders of Object Recognition and What They Tell Us About Normal Vision* (MIT Press, Cambridge, MA, 1990).
4. Gomori, A. & Hawryluk, G. *Neurology* **34**, 947–950 (1984).
5. Albert, M., Soffer, D., Silverberg, R. & Reches, A. *Neurology* **29**, 870–879 (1979).
6. Mack, J. & Bollen, F. *Neuropsychologia* **15**, 345–348 (1977).
7. Bub, D., Black, S., Hampson, E. & Kertesz, A. *Cogn. Neuropsychol.* **5**, 27–66 (1988).
8. Beauvois, M. *Phil. Trans. R. Soc. Lond. B* **298**, 35–47 (1982).
9. Warrington, E. & McCarthy, R. *Neuropsychologia* **32**, 1465–1473 (1994).
10. Warrington, E. & Shallice, T. *Brain* **107**, 829–853 (1984).
11. McCarthy, R. & Warrington, E. *Nature* **334**, 428–430 (1988).
12. Glaser, W. *Cognition* **42**, 61–105 (1992).
13. Caramazza, A., Hillis, A., Rapp, B. & Romani, C. *Cogn. Neuropsychol.* **7**, 161–189 (1990).
14. Butterworth, B., Howard, D. & McLoughlin, P. *Neuropsychologia* **22**, 409–426 (1984).
15. Howard, D. & Patterson, K. *Pyramids and Palm Trees: a Test of Semantic Access from Pictures and Words*. (Thames Valley, Bury St Edmunds, 1992).
16. te Linde, J. J. *Exp. Psychol. Learn. Mem. Cogn.* **8**, 584–598 (1982).
17. Martin, A., Wiggs, C., Ungerleider, L. & Haxby, J. *Nature* **379**, 649–652 (1996).
18. Demonet, J.-F. et al. *Brain* **115**, 1753–1768 (1992).
19. Price, C., Moore, C. & Frackowiak, R. S. J. *Neuroimage* **3**, 40–52 (1996).
20. Martin, A., Haxby, J., Lalonde, F., Wiggs, C. & Ungerleider, L. *Science* **270**, 102–105 (1995).
21. Price, C., Wise, R. & Frackowiak, R. S. J. *Cereb. Cort.* **6**, 62–70 (1996).
22. Damasio, H., Grabowski, T., Tranel, D., Hichwa, R. & Damasio, A. *Nature* **380**, 499–505 (1996).
23. Petersen, S., Fox, P., Snyder, A. & Raichle, M. *Nature* **331**, 585–589 (1988).
24. Ridoch, J. & Humphreys, G. *Cogn. Neuropsychol.* **4**, 131–185 (1987).
25. Coltheart, V., Patterson, K. & Leahy, J. Q. J. *Exp. Psychol.* **47A**, 917–955 (1994).
26. Mishkin, M., Ungerleider, L. & Macko, K. *Trends Neurosci.* **6**, 414–417 (1983).
27. Wilson, F., Scaldache, S. O. & Goldman-Rakic, P. *Science* **260**, 1955–1958 (1993).
28. Snodgrass, J. & Vanderwart, M. J. *Exp. Psychol. Hum. Learn. Mem.* **6**, 174–215 (1980).
29. Friston, K. et al. *Hum. Brain Mapp.* **2**, 189–210 (1995).
30. Friston, K. et al. *Hum. Brain Mapp.* **2**, 165–189 (1995).

ACKNOWLEDGEMENTS. We thank the volunteers who participated in this study. R.V. is supported by the Belgian National Research Council.

CORRESPONDENCE and requests for materials should be addressed to C.P. (e-mail: cprice@fil.ion.ucl.ac.uk).

Contribution of human hippocampal region to novelty detection

Robert T. Knight

Department of Neurology, Center for Neuroscience, University of California, Davis, Veterans Medical Center, 150 Muir Road, Martinez, California 94553, USA

THE ability to respond to unexpected stimuli (the ‘orienting response’) is a fundamental characteristic of mammalian behaviour¹, but the brain mechanisms by which novelty is detected remain poorly defined. Electrophysiological recordings of scalp and intracranial event-related potentials (ERPs) have shown that novel stimuli activate a distributed network involving prefrontal and posterior association cortex^{2–6}. In addition, ERP^{7,8} and single-neuron^{9,10} recordings, as well as neuroimaging¹¹ and modelling¹² studies, have suggested that temporal cortical regions, including the hippocampus, are also involved. To examine further the role of the medial temporal lobe in novelty processing, I measured physiological responses to novel auditory and tactile stimuli in patients with damage to the posterior hippocampal region. In normal control subjects, unexpected novel stimuli produce a characteristic ERP signal, accompanied by an autonomic skin response. Both responses are reduced in hippocampal lesion patients, whereas the response to expected control stimuli is unaffected. Thus the hippocampal region, in addition to its known role in memory formation, is an essential component of the distributed limbic–cortical network that detects and responds to novel stimuli.

Distinctive ERPs are recorded from the scalp of humans in

response to stimuli that stand out from others in the perceptual stream. Such events may be salient either because they are relevant signals requiring a response or because they are novel and surprising. The ERP scalp signature for these two types of signals exhibits different latencies and topographies. Novel stimuli or stimuli sufficiently deviant from the ongoing sensory environment trigger a positive brain potential maximal over fronto-central regions (250–550 ms; P3a) which is proposed to be a central marker of the orienting response^{13,14}. Voluntary detection of a task-relevant target stimulus elicits a longer latency positivity (300–600 ms; P3b) which is maximal over parietal sites. Target stimuli may also elicit overlapping, early latency P3a activity dependent on the degree of target deviancy from background events.

I performed experiments in which target and novel stimuli (tones, complex tones; finger taps, wrist shocks), embedded in trains of repetitive background stimuli, were delivered randomly to subjects engaged in detection tasks. Target stimuli differed only slightly from the background stimuli, whereas novel stimuli differed substantially in order to be highly distinctive. Subjects were patients with unilateral brain lesions centred in the posterior hippocampal region (Fig. 1) and age-matched controls. They sat in a sound-attenuated booth and pressed a button upon detection of a designated target stimulus during separate auditory and somatosensory testing sessions. ERPs were recorded from scalp electrodes to background, target and novel stimuli.

In controls, target stimuli elicited a prominent parietal–maximal P3b in both modalities and a small, shorter latency fronto-central P3a. Novel stimuli in both modalities generated enhanced amplitude P3a potentials which had more fronto-central scalp distributions ($P < 0.001$ for target versus novel scalp topography; Figs 2 and 3) and shorter latencies than the target P3bs (novels at Fz: auditory, 365 ms; somatosensory, 354 ms; targets at Pz: auditory, 431 ms, somatosensory, 410 ms; $P < 0.002$ for both modalities).

Hippocampal lesions had differential effects on P3b and P3a

responses. Posterior scalp P3b activity to target stimuli was not affected by hippocampal damage (Fig. 2a, b), in accord with reports of normal P3b amplitudes in patients with hippocampal damage due to anoxia-induced CA1 damage, infiltrative tumour or herpes simplex encephalitis^{15,16}. Coupled with lesion^{3,4} and intracranial^{5,7} data, these findings indicate that the posterior scalp P3b measures cortical neural activity generated during controlled attention to a task-relevant events.

In contrast to the lack of effect on the P3b, hippocampal damage resulted in widespread reductions of the novelty P3a, with decrements being most pronounced over prefrontal regions ($P < 0.001$ for both modalities; Figs 2a, b and 3). Comparable novelty response reductions were observed for left and right hippocampal lesions and for stimuli presented ipsilaterally or contralaterally to the lesions. The smaller fronto-central P3a's generated to target stimuli were similarly reduced by hippocampal damage ($P < 0.05$ for both modalities).

The amplitude and habituation characteristics of peripheral autonomic sympathetic skin responses (SSRs) to random wrist shock were also altered by hippocampal damage. SSR amplitude was reduced and there was a flat habituation curve in the hippocampal subjects (Fig. 2c, d; amplitude and habituation both $P < 0.005$). Reductions were observed for orienting stimuli presented ipsilaterally and contralaterally to the lesion. These results reveal that both peripheral and central orienting responses are impaired by mesial temporal damage. Previous studies in normals have reported correlations between the amplitudes of the P3a and the SSR, indicating that these potentials measure combined central and peripheral orienting responses to novel or sufficiently deviant stimuli⁷.

Novelty P3a potentials are reduced in amplitude after focal damage in either dorsolateral prefrontal or posterior association cortex²⁻⁴, and intracranial electrodes record novelty-related field potentials in several brain regions, including prefrontal and posterior association cortex, in addition to cingulate and limbic areas⁵⁻⁸. This suggests that widely distributed circuits in the human brain are activated by stimulus novelty. My findings provide evidence that the human hippocampal region is a critical element of this novelty detection network. Neural circuits dependent on the hippocampal region appear to maintain a template of the recent past¹⁷ for comparison of incoming sensory stimuli. Deviation from this template activates a distributed limbic-neocortical orienting system which facilitates behavioural response to and memory storage of discrete novel events.

Behavioural studies in the 1930s (ref. 18) and subsequent behavioural-electrophysiological data¹⁹ support the notion that stimulus novelty enhances recall. Lesion and intracranial ERP data document multimodal bilateral activation of prefrontal, cingulate, temporal-parietal and hippocampal regions during processing of novel events²⁻⁸. Positron emission tomography

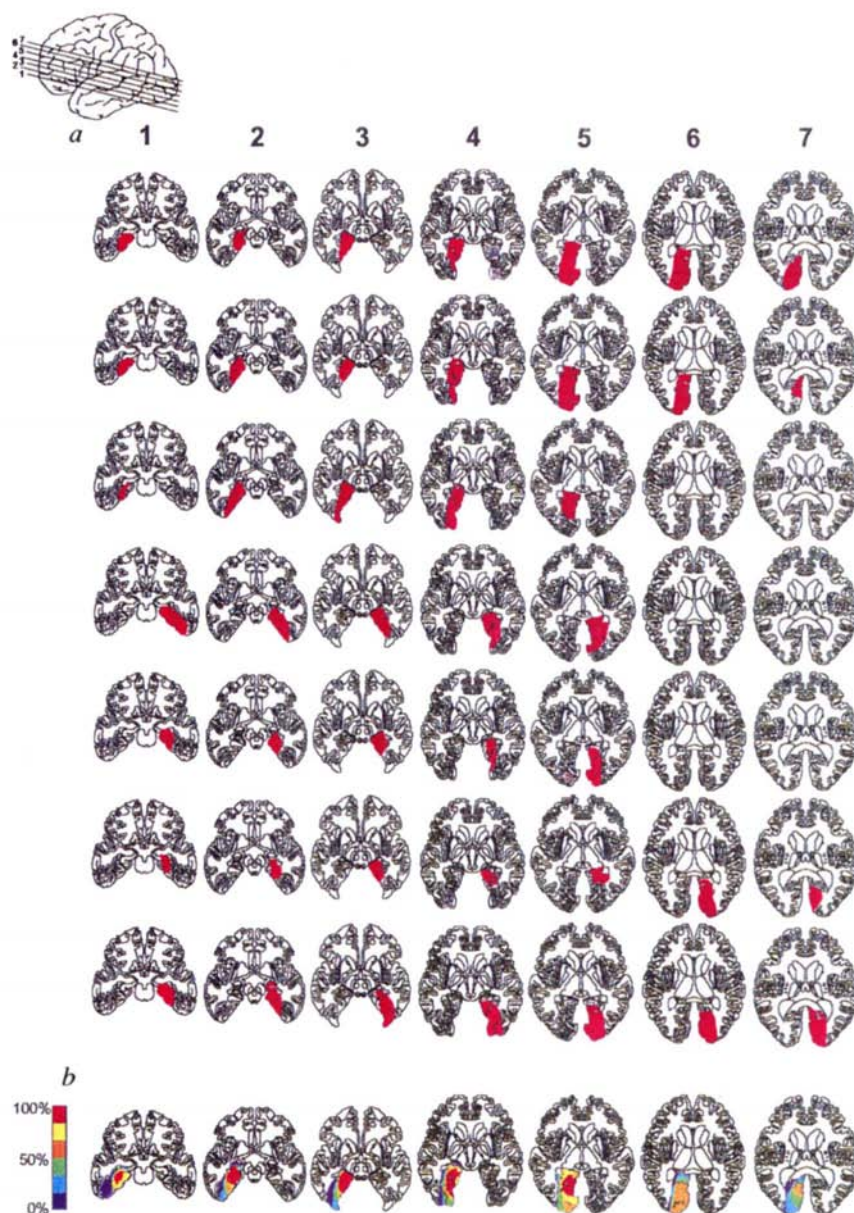


FIG. 1 a, Lesion reconstructions are shown for individual hippocampal patients. All damage was due to single infarction in the distribution of the posterior cerebral artery including the branch of Uchimura perfusing the posterior hippocampus. Lesions are transcribed onto axial templates using 5-mm cuts. Each row shows the extent of damage in an individual patient. b, The group-averaged lesion extent is shown. Both left and right hippocampal lesions are reflected onto the left side for averaging. All lesions overlapped in the posterior hippocampal region (red; axial cuts 1–5). The average tissue loss was 23.9 cm³ per patient. The volume of common tissue loss (100% lesion overlap; red) was 11.4 cm³.

(PET) studies provide further support for the idea that stimulus novelty activates distributed brain regions engaged in memory storage, reporting predominantly bilateral posterior cortical and right hippocampal involvement¹¹. This divergence may be due to the fact that PET integrates peaks of activation over sustained temporal epochs, whereas ERPs measure phasic neural activity specific to the novel stimulus.

Single-unit data in monkeys reveals that novelty-related cells in inferior temporal cortex fire within 100 ms (ref. 10), suggesting that stimulus processing antecedent to P3a generation may contribute to novelty-related activation in humans. Inter- and intra-hemispheric cortico-cortical and cortico-limbic connections^{20,21} engaged during attention and memory performance²² may

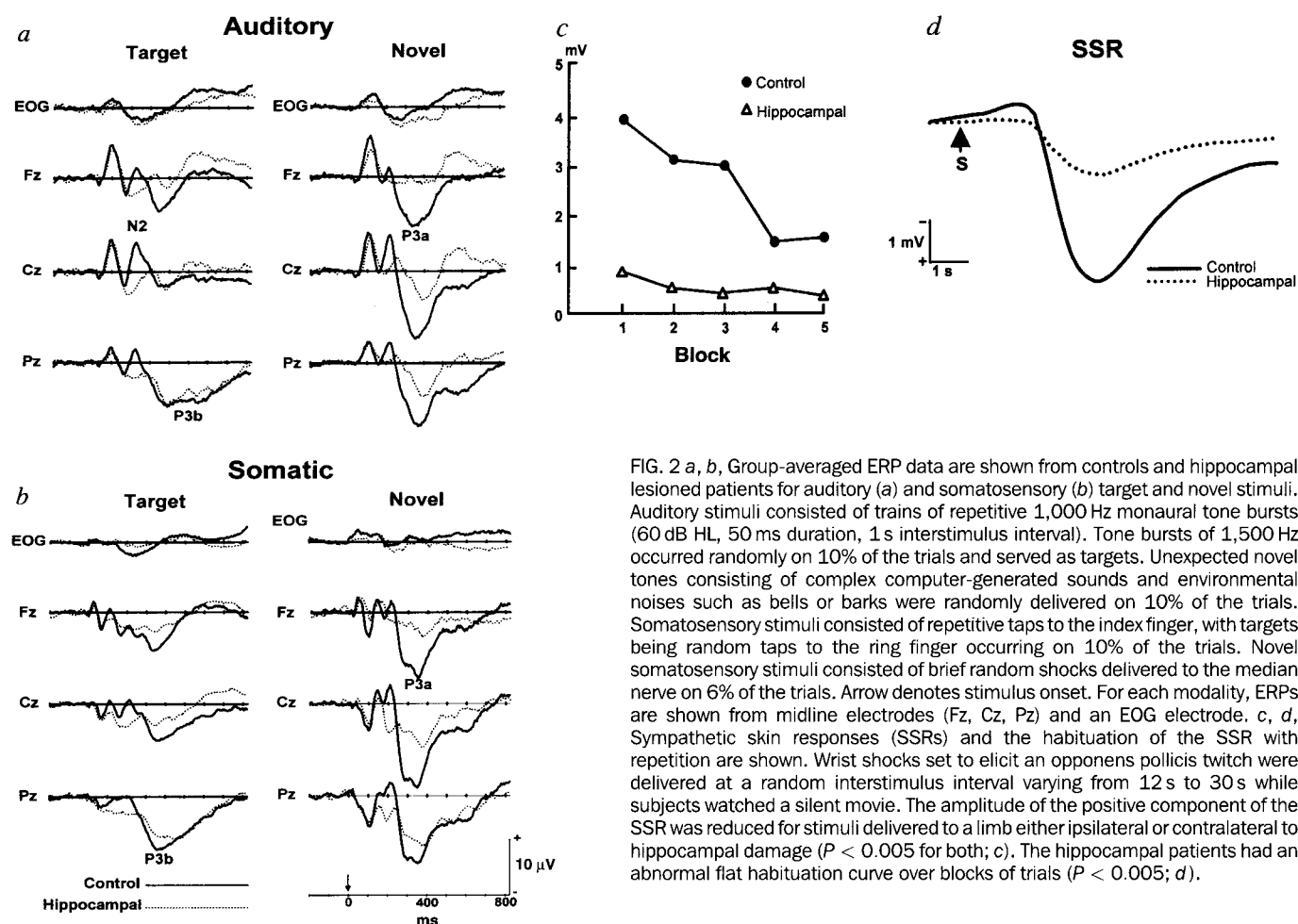


FIG. 2 *a, b*, Group-averaged ERP data are shown from controls and hippocampal lesioned patients for auditory (*a*) and somatosensory (*b*) target and novel stimuli. Auditory stimuli consisted of trains of repetitive 1,000 Hz monaural tone bursts (60 dB HL, 50 ms duration, 1 s interstimulus interval). Tone bursts of 1,500 Hz occurred randomly on 10% of the trials and served as targets. Unexpected novel tones consisting of complex computer-generated sounds and environmental noises such as bells or barks were randomly delivered on 10% of the trials. Somatosensory stimuli consisted of repetitive taps to the index finger, with targets being random taps to the ring finger occurring on 10% of the trials. Novel somatosensory stimuli consisted of brief random shocks delivered to the median nerve on 6% of the trials. Arrow denotes stimulus onset. For each modality, ERPs are shown from midline electrodes (Fz, Cz, Pz) and an EOG electrode. *c, d*, Sympathetic skin responses (SSRs) and the habituation of the SSR with repetition are shown. Wrist shocks set to elicit an opponens pollicis twitch were delivered at a random interstimulus interval varying from 12 s to 30 s while subjects watched a silent movie. The amplitude of the positive component of the SSR was reduced for stimuli delivered to a limb either ipsilateral or contralateral to hippocampal damage ($P < 0.005$ for both; *c*). The hippocampal patients had an abnormal flat habituation curve over blocks of trials ($P < 0.005$; *d*).

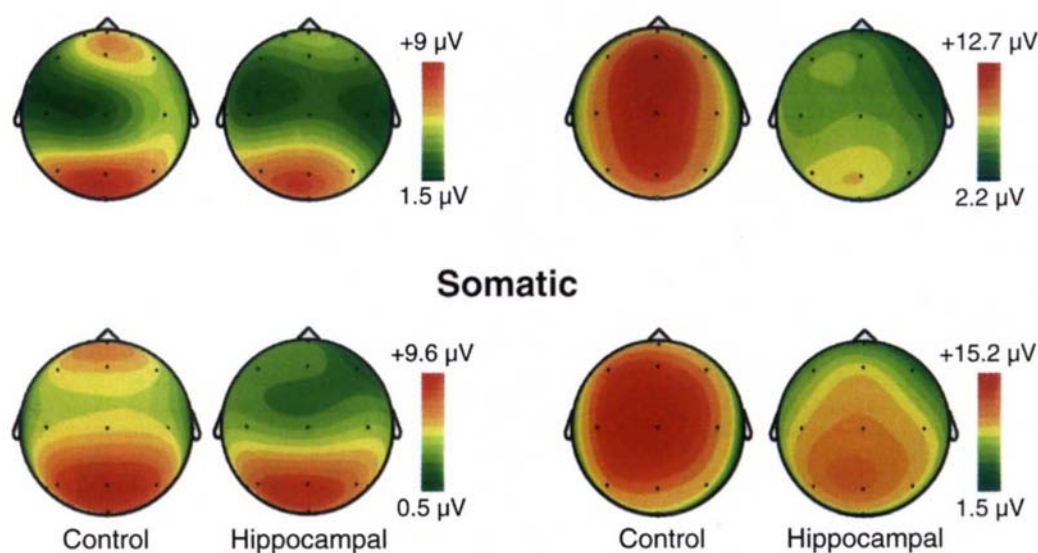


FIG. 3 Spherical spline topographical maps for the P3a novelty responses in control and hippocampal patients in both the auditory and somatosensory modalities. I used 18 electrodes to produce the spline maps. The voltage

maps show the increase in fronto-central P3 activity for novel versus target stimuli in controls, and the pronounced novelty P3a reductions in the hippocampal patients.

provide the anatomical substrate for novelty-related limbic-cortical activation, with hippocampal-hypothalamic pathways subserving the peripheral autonomic orienting response²³.

P3a-like potentials have been reported in several mammalian species, suggesting that this hippocampal-neocortical network may provide a core substrate for orientation to novelty. The findings described here provide evidence that the hippocampal region is a central component of a distributed neural network engaged during orientation to perturbations in the sensory stream. □

Methods

Patient selection. Damage caused by infarction of the posterior cerebral artery (PCA) centred in the posterior hippocampal region, including hippocampus proper, dentate gyrus, subiculum, parahippocampal gyrus, entorhinal cortex and fornix in all patients (see ref. 24 for parcellation of anatomical subregions of the hippocampal region). Variable amounts of damage to adjacent fusiform, lingual and calcarine gyri existed in individual patients. Resting PET scanning using [¹⁸F]fluorodeoxyglucose in four of the patients revealed additional hypometabolism in anterior mesial temporal regions not structurally damaged by infarction²⁵. There were 3 left- and 4 right-lesioned patients (64 ± 6 years). All patients were tested at least 6 months post-stroke and damage was estimated from magnetic resonance imaging (MRI) or computed tomography (CT) scanning. Controls were matched for age (61 ± 8 years) and education to the patients and were free from neurological or psychiatric disease. Left hippocampal patients had predominantly verbal deficits and right hippocampal patients had non-verbal memory deficits. Two patients with splenial involvement (top 2 cases in Fig. 1) had the behavioural syndrome of alexia without agraphia. Similar neuropsychological profiles have been reported in patients with PCA infarcts^{26,27}. The research was approved by the Human Subjects Review Committees of the University of California, Davis and the Veterans Administration Medical Research Service.

Electrophysiological recording.

ERPs. Electroencephalograms (0.1–100 Hz; 256 Hz digitization rate) were recorded from Fp1, Fp2, F3, F4, Fz, T3, T4, T5, T6, C3, C4, Cz, P3, P4, Pz, O1, Oz, O2 in the international 10–20 system. Electro-oculography (EOG) electrodes were located vertical and horizontal to the eye. Electrodes were referred to a balanced non-cephalic sterno-vertebral reference. Trials with excessive ocular or electromyogram (EMG) artefact were rejected. Sensory thresholds and early latency potentials did not differ between groups. Subjects pressed a button to detected targets. Detection accuracy and reaction times were comparable between groups. Novel stimuli preceding a target did not affect detection reaction time in either group. Novel-target interstimulus intervals of <400 ms prolong response to targets²⁸ and interstimulus intervals were kept at 1,000 ms to avoid overlapping ERP activity. Replication blocks with 40 target and novel stimuli were obtained for each ear and limb. Order of stimulation was counterbalanced across subjects and lesion location. Comparable results were obtained independently of ear or limb stimulation in both groups and the data were collapsed across blocks for presentation. Peak and mean amplitude for P3bs to targets (300–600 ms), P3as to novels (250–550 ms) and the N2 to novels and targets (180–260 ms) were quantified relative to a 200-ms prestimulus baseline. Latency was recorded relative to stimulus onset. Mean amplitude was measured in a 50-ms window centred on each subject's peak latency. Voltage amplitude s.e.m. ranged from 1.1–1.9 μ V at midline sites. Scalp topographies were statistically analysed after voltage normalization. The use of 18 electrodes may have resulted in minor voltage-map distortions at peripheral scalp sites. All measurements were subjected to within and between groups ANOVA with Geisser–Greenhouse corrections. The N2 response, associated with sensorimotor integration²⁹, was comparably reduced by both target and novel stimuli (hippocampal patients: $P < 0.05$ for both modalities).

SSRs. Skin potentials were recorded from Ag–AgCl electrodes affixed to the palm and dorsum of the hand³⁰. Signals were preamplified with a J&J IG-3 preamp and the analogue output was amplified (60 $\times$) through a Grass P511 AC amplifier (0.01–30 Hz), digitized (64 Hz) and stored for off-line analysis and averaging. Stimuli consisted of brief electrical shocks (0.02 ms duration) delivered from a Grass S88 constant current stimulator to electrodes placed over the median nerve at the wrist. Voltage was adjusted to obtain a just visible twitch of the opponens pollicis muscle. A total of 60 random shocks were delivered randomly ipsilaterally or contralaterally to hippocampal damage. Data from the limb contralateral to the lesion is shown in Fig. 2c (1 mV corresponds to 0.025 μ s). The habituation curve (Fig. 2d) includes averaged SSRs from sequential blocks of 10 trials, with amplifier blocking responses excluded.

- Halgren, E. et al. *Electroenceph. Clin. Neurophysiol.* **94**, 191–220 (1995).
- Baudena, P., Halgren, E., Heit, G. & Clarke, J. M. *Electroenceph. Clin. Neurophysiol.* **94**, 251–264 (1995).
- Halgren, E. & Marinkovic, K. in *The Cognitive Neurosciences* (ed. Gazzaniga, M. S.) 1137–1151 (MIT Press, Cambridge, MA, 1995).
- Scabini, D. & McCarthy, G. *Soc. Neuroscience (abstr.)* **19**, 564 (1993).
- Rolls, E. T., Cahusac, P. M. B., Feigenbaum, J. D. & Miyashita, Y. *Exp. Brain Res.* **93**, 299–306 (1993).
- Miller, E. K., Li, L. & Desimone, R. *Science* **254**, 1377–1380 (1991).
- Tulving, E., Markowitsch, H. J., Craik, F. I. M., Habib, R. & Houle, S. *Cerebral Cortex* **6**, 71–79 (1996).
- Metcalfe, J. *Psychol. Rev.* **100**, 3–22 (1993).
- Squires, N. K., Squires, K. C. & Hillyard, S. A. *Electroenceph. Clin. Neurophysiol.* **63**, 387–401 (1975).
- Courchesne, E., Hillyard, S. A. & Galambos, R. *Electroenceph. Clin. Neurophysiol.* **39**, 131–143 (1975).
- Polich, J. & Squires, L. R. *Electroenceph. Clin. Neurophysiol.* **86**, 408–417 (1993).
- Onofri, M. et al. *Neurology* **42**, 1762–1767 (1992).
- Treves, A. & Rolls, E. T. *Hippocampus* **4**, 374–391 (1994).
- von Restorff, H. *Psychol. Forsch.* **18**, 299–342 (1933).
- Karis, D., Fabiani, M. & Donchin, E. *Cogn. Psychol.* **16**, 177–216 (1984).
- Goldman-Rakic, P. S., Selemon, L. D. & Schwartz, M. L. *Neuroscience* **12**, 719–743 (1984).
- Van Hoesen, G. W., Morecraft, R. J. & Semendeferi, K. in *Neuropsychiatry* (eds Fogel, B. S., Schiffer, R. B. & Rao, S. M.) 113–143 (Williams and Wilkins, Baltimore, 1996).
- Friedman, H. R. & Goldman-Rakic, P. S. *J. Neurosci.* **14**, 2775–2788 (1994).
- Risold, P. Y. & Swanson, L. W. *Science* **272**, 1484–1486 (1996).
- Eichenbaum, H., Otto, T. & Cohen, N. *Behav. Brain Sci.* **17**, 449–518 (1994).
- Richardson, B. C., Eberling, J. C., Knight, R. T. & Jagust, W. J. *Neurology* **42**, 170 (1992).
- De Renzi, E., Zambolin, A. & Crisi, G. *Brain* **110**, 1099–1116 (1987).
- von Cramon, D. Y., Hebel, N. & Schuri, U. *Brain* **111**, 1061–1077 (1988).
- Woods, D. L., Knight, R. T. & Scabini, D. *Cogn. Brain Res.* **1**, 227–240 (1993).
- Renault, B., Ragot, R., Lesevre, N. & Remond, A. *Science* **215**, 1413–1415 (1982).
- Gutrecht, J. A. J. *Clin. Neurophysiol.* **11**, 519–524 (1994).

ACKNOWLEDGEMENTS. This work was supported by a grant from the NINDS. I thank C. Clayworth for technical assistance in all phases of the project.

CORRESPONDENCE and requests for material should be addressed to the author (e-mail: rtknight@ucdavis.edu).

Regulation of rhythmic movements by purinergic neurotransmitters in frog embryos

Nicholas Dale* & Deborah Gilday†

* School of Biological and Medical Sciences, Bute Medical Building, University of St Andrews, St Andrews, KY16 9RU, UK

† School of Biological Sciences, University of Bristol, Woodland Road, Bristol, BS8 1UG, UK

MANY rhythmic motor behaviours, including swimming¹, walking², scratching³, swallowing⁴, micturition⁵ and sexual climax⁶, are episodic: even in the absence of sensory inputs they exhibit a gradual run-down in frequency before spontaneously terminating. We have investigated whether the purinergic transmitters, ATP and adenosine, control run-down of swimming in the *Xenopus* embryo⁷. By using specific agonists and antagonists for the purinergic receptors, we have shown that ATP (or a related substance) is released during swimming and activates P2y receptors to reduce voltage-gated K⁺ currents and cause an increase in the excitability of the spinal motor circuits. Adenosine is also produced during motor activity, possibly through the actions of ectonucleotidases. The activation by adenosine of P1 receptors reduces the voltage-gated Ca²⁺ currents, lowers excitability of the motor circuits, and so opposes the actions of ATP. A gradually changing balance between ATP and adenosine therefore seems to underlie the run-down of the motor pattern for swimming in *Xenopus*. We believe this to be the first time that ATP and adenosine have been found to be involved in motor pattern generation. The antagonistic interplay between these two transmitters may offer a general feedback mechanism that underlies run-down of all episodic motor patterns in vertebrates.

The actions of purinergic agonists and antagonists on the length of swimming episodes in the stage 37/38 *Xenopus* embryo were studied. The effects of ATP (10–100 μ M) were variable, but there

Received 12 June; accepted 2 August 1996.

- Sokolov, E. N. *Annu. Rev. Physiol.* **25**, 545–580 (1963).
- Knight, R. T. *Electroenceph. Clin. Neurophysiol.* **59**, 9–20 (1984).
- Knight, R. T., Scabini, D., Woods, D. L. & Clayworth, C. C. *Brain Res.* **502**, 109–116 (1989).
- Yamaguchi, S. & Knight, R. T. *J. Neurosci.* **11**, 2039–2054 (1991).

was overall a slight but significant reduction in the length of episodes (Fig. 1a, b). To eliminate any possible confounding effects of the breakdown of ATP to adenosine by ectonucleotidases, 10 μ M ATP was applied in the presence of an adenosine receptor antagonist, 8-phenyltheophylline (8PT; 2.5 μ M)⁸. Under these conditions, ATP consistently and reliably produced great lengthening of swimming episodes (Fig. 1a, b).

To examine the possible role of P2 receptors, the antagonists phosphate-6-azophenyl-2',4'-disulphonate (PPADS)⁹ and suramin¹⁰ were applied to spinal embryos. Both PPADS and suramin at 10 μ M greatly reduced episode length (Fig. 1c, d). Because the neural circuit that controls swimming uses glutamate-mediated synaptic excitation and glycinergic reciprocal inhibition, it was important to eliminate possible nonspecific actions of PPADS and suramin on these transmitters. We therefore tested whether PPADS and suramin blocked inward currents evoked by glutamate, NMDA (*N*-methyl-D-aspartate) and glycine in acutely isolated *Xenopus* spinal neurons. Neither PPADS nor suramin affected responses to 100 μ M glutamate (Fig. 1e; mean change $\pm$ s.e.m. $-3.2 \pm 1.9\%$, $n = 10$, and $4.5 \pm 5.6\%$, $n = 8$, respectively). Similarly, PPADS had no effect on responses to 100 μ M NMDA (Fig. 1e; $1.1 \pm 6.8\%$, $n = 9$) or 100 μ M glycine (Fig. 1e; $0.8 \pm 0.8\%$, $n = 4$). However, suramin significantly reduced the NMDA current (Fig. 1e; $-57.4 \pm 3.9\%$, $n = 7$, $P < 0.02$, paired sample *t*-test) and the response to glycine (Fig. 1e; $-23.0 \pm 5.7\%$, $n = 5$, $P < 0.05$, paired sample *t*-test). In additional tests, 10 μ M PPADS had no effect on the voltage-gated Na^+ , Ca^{2+} or K^+ currents. Because PPADS did not block receptors for glutamate, NMDA or glycine, its effects on swimming can best be explained by a direct action on P2 receptors. We therefore conclude that P2 receptors are activated during swimming by an endogenous ligand, presumably ATP, and act to increase the excitability of motor circuits. The effects of suramin may be partly mediated by its modest block of NMDA and glycine receptors.

We next tested whether adenosine might also modulate the locomotor circuitry. Adenosine (10–100 μ M) caused a drastic shortening of episode length (Fig. 2a, d), whereas the antagonist 8PT greatly increased the duration of swimming episodes (Fig. 2b, d), suggesting that endogenous adenosine decreases the excitability of the motor circuits. We tested the pharmacological specificity of 8PT against the transmitters that mediate fast-synaptic transmission in the swimming circuit. We found that 2.5 μ M 8PT had no effect on the inward currents evoked by: 100 μ M glutamate (Fig. 2e; $-1.5 \pm 2.4\%$, $n = 5$); 200 μ M NMDA (Fig. 2f; $0.3 \pm 5.0\%$, $n = 8$); or 100 μ M glycine (Fig. 2f; $1.2 \pm 1.4\%$, $n = 5$). The lengthening of swimming episodes caused by 8PT is probably due to a specific action on adenosine receptors.

Rather than being released from nerve terminals, adenosine might be produced from ATP through the actions of ectonucleotidases. Application of a blocker of the 5'-nucleotidase α,β -methylene-ADP¹¹ (AMPCP; Fig. 2c, d), also increased the length of swimming episodes. This, together with the observation that ATP is much more effective at lengthening swimming episodes when applied in the presence of an adenosine antagonist, suggests that adenosine is produced from the breakdown of ATP.

To investigate the possible cellular mechanisms underlying the effects of ATP and adenosine, intracellular recordings were made from motor neurons in the spinal cord. Neither adenosine nor ATP caused a change in the resting membrane potential or input resistance (Fig. 3a, b). In addition, ATP and adenosine caused no significant change in the excitatory (mediated by AMPA and NMDA receptors) or inhibitory (mediated by glycine receptors) synaptic drive to motor neurons during swimming (Fig. 3c–e). To test further whether these transmitters affected synaptic transmission, axons on the outside of the spinal cord were stimulated to evoke postsynaptic potentials in motor neurons. Inhibitory glycinergic and excitatory glutamatergic postsynaptic potentials were pharmacologically isolated, and the effects of ATP and adenosine on synaptic transmission studied. Neither adenosine nor ATP

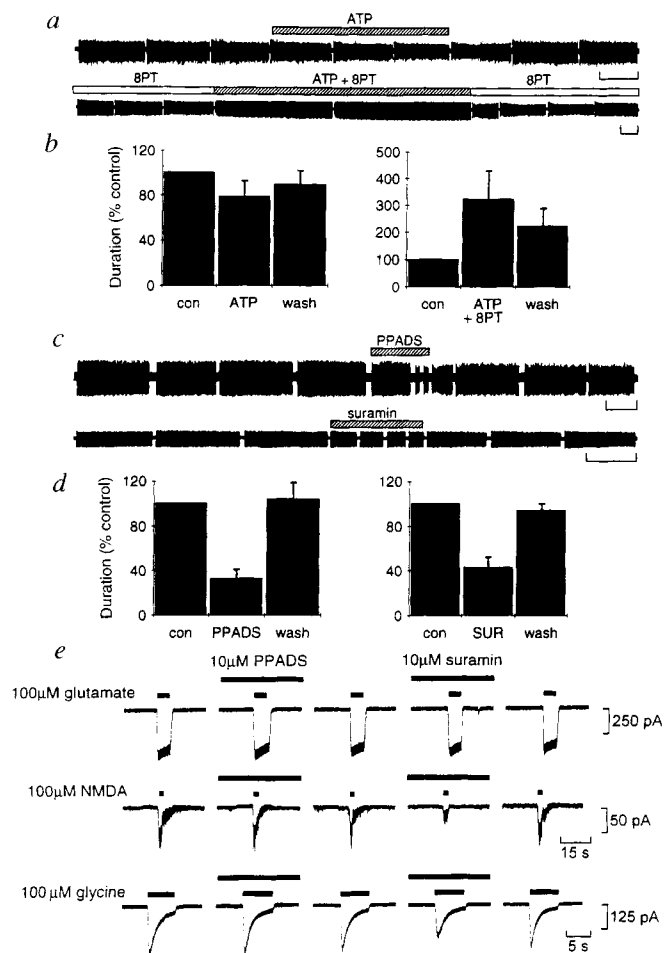


FIG. 1 ATP increases the excitability of the spinal circuitry. **a**, Effects on swimming of 10 μ M ATP applied alone and with 2.5 μ M 8PT, an adenosine receptor antagonist. **b**, Summary histograms for ATP and ATP + 8PT (7 embryos; Wilcoxon matched-pairs signed-rank test: $P < 0.05$ and $P < 0.02$, respectively). **c**, Effects of PPADS (10 μ M) and suramin (10 μ M) on the length of swimming episodes. Scale bars, 2 min in **a** and **c**. **d**, Summary histograms for PPADS in 9 embryos and suramin in 5 embryos. **e**, Responses to glutamate, NMDA and glycine recorded in isolated neurons during application of PPADS and suramin.

significantly affected the amplitude of the glycinergic inhibitory postsynaptic potentials (IPSPs) or glutamatergic excitatory postsynaptic potentials (EPSPs) (Fig. 3f, g). The amplitude of the evoked IPSP was -10.5 ± 2.0 mV and -10.1 ± 1.8 mV ($n = 12$) in the control and 100 μ M adenosine, respectively. Similarly, 100 μ M ATP applied either on its own ($n = 9$) or in combination with 2.5 μ M 8PT ($n = 4$) had no significant effect on the IPSP (amplitudes, -11.5 ± 1.9 mV and -11.3 ± 2.0 mV ($n = 13$) in control and ATP, respectively). For the evoked EPSP, the mean amplitude was 10.5 ± 1.6 mV and 9.8 ± 1.7 mV ($n = 9$) in the control and 100 μ M adenosine, respectively; and 9.6 ± 1.5 mV and 9.5 ± 1.4 mV ($n = 5$) in the control and 100 μ M ATP plus 2.5 μ M 8PT, respectively.

We next tested whether these purinergic transmitters could modulate voltage-gated currents by performing whole-cell patch-clamp recordings from motor neurons and premotor interneurons (known to participate in the motor circuit) acutely isolated from the embryo spinal cord. ATP reduced the voltage-gated K^+ currents reversibly ($-13.8 \pm 1.1\%$, $n = 20$, $P < 0.001$, paired sample *t*-test; Fig. 4a) but did not cause a change in the leak or Ca^{2+} currents. In contrast, adenosine reversibly reduced the voltage-gated Ca^{2+} currents ($-15.7 \pm 2.3\%$, $n = 11$, $P < 0.001$; Fig. 4b) but had no effect on the leak or K^+ currents. Because

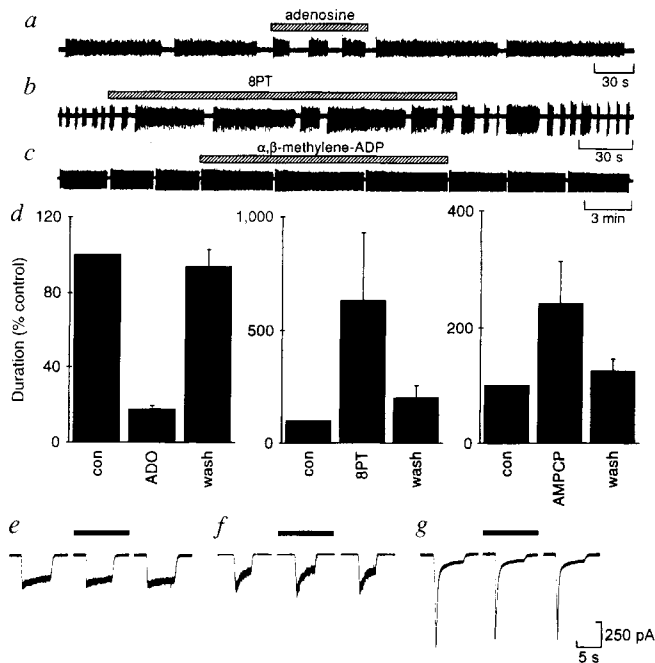


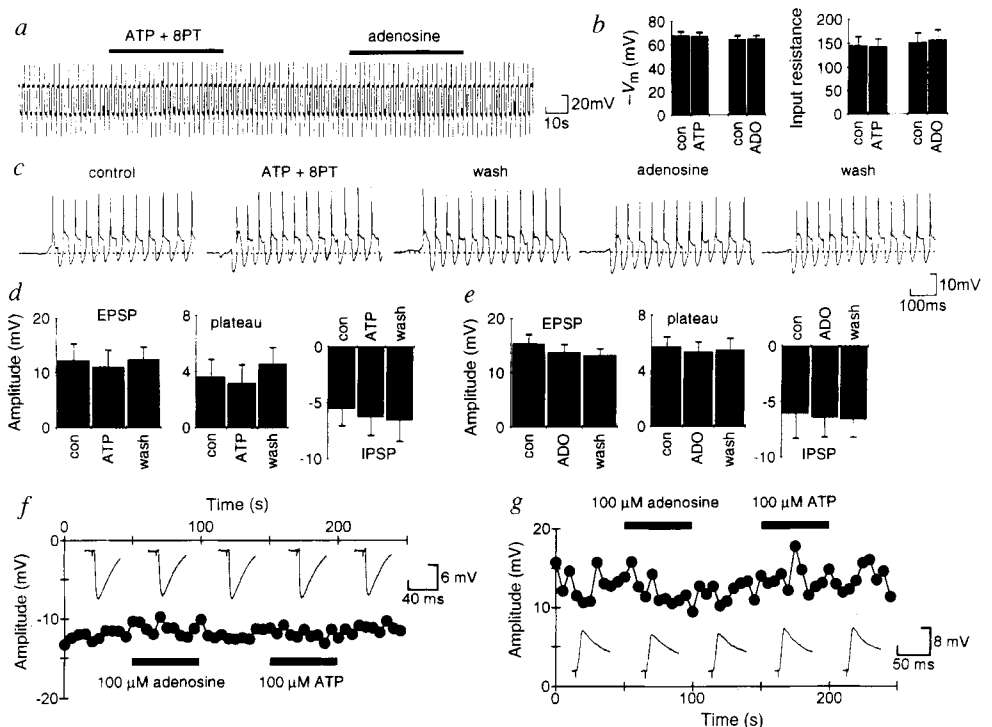
FIG. 2 Endogenous production of adenosine reduces the excitability of spinal circuits. *a*, Effect of 100 μ M adenosine, *b*, 2.5 μ M 8PT and *c*, 100 μ M α, β -methylene-ADP, on length of swimming episodes. *d*, Summary histograms for 100 μ M adenosine (ADO) in 4 embryos ($P < 0.03$, paired-sample *t*-test), 2.5 μ M 8PT in 12 embryos, and 100 μ M α, β -methylene-ADP (AMPCP) in 9 embryos ($P < 0.01$ for both, Wilcoxon matched-samples signed-rank test). 2.5 μ M 8PT (applied during bar) had no effect on responses to 100 μ M glutamate (*e*), 200 μ M NMDA (*f*) or 100 μ M glycine (*g*).

neither ATP nor adenosine affected synaptic transmission, this modulation of currents may occur in the soma rather than in the synaptic terminals.

Reduction of voltage-gated K^+ currents by ATP will reduce neuronal thresholds, enhance excitability of the locomotor circuitry, and thus allow generation of longer episodes of swimming. By reducing the voltage-gated Ca^{2+} currents, adenosine is likely to have the opposite effect and cause shortening of swimming episodes (Fig. 4c). Can the contradictory actions of these two transmitters be reconciled? In many systems the 5'-nucleotidase that converts AMP to adenosine is inhibited by ATP and ADP¹¹,

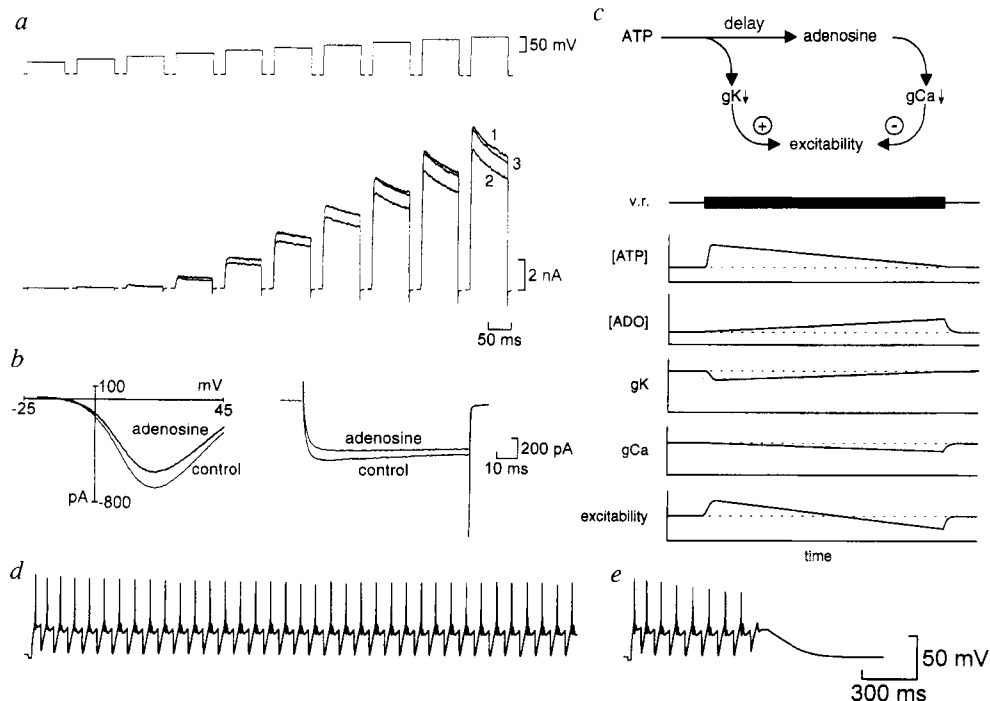
which delays the appearance of adenosine with respect to the release of ATP. If this were to occur in the embryo spinal cord, the levels of ATP and adenosine would slowly change during swimming (Fig. 4c). According to our hypothesis, ATP levels will be high and those for adenosine low early in a swimming episode. Hence the Ca^{2+} currents will be maximal, the K^+ currents will be at their lowest, and the locomotor circuitry will be in a state of high excitability (Fig. 4c). As swimming proceeds, adenosine concentrations slowly increase, thereby reducing the Ca^{2+} currents. Meanwhile, the level of ATP will fall as it is converted to adenosine and as the frequency of swimming declines, leading

FIG. 3 Neither adenosine nor ATP directly gate an ion channel or change the synaptic drive to motor neurons during swimming. *a*, *b*, Effects of ATP and adenosine on resting potential ($n = 18$ and $n = 16$, respectively) and input resistance in spinal motor neurons (measured in $M\Omega$; $n = 8$ and $n = 7$, respectively). In 6 neurons ATP was applied in presence of 2.5 μ M 8PT. *c*, Intracellular recording from spinal motor neuron during application of 100 μ M ATP plus 2.5 μ M 8PT and 100 μ M adenosine. In this example, adenosine caused a small decrease in the synaptic excitation. Summary results for ATP (*d*; $n = 7$, applied in presence of 2.5 μ M 8PT in 4 cases) and adenosine (*e*; $n = 8$). Although the size of the EPSP in adenosine was significantly reduced with respect to control ($P < 0.05$, Wilcoxon matched-pairs signed-rank test), this effect did not reverse on wash, and when all three treatments (control, drug and wash) were analysed by a Kruskal-Wallis single-factor analysis of variance (ANOVA) there was no significant change in the size of the synaptic drive. Neither 100 μ M ATP nor 100 μ M adenosine affect amplitude of evoked IPSPs (*f*) or EPSPs (*g*) recorded in motor neurons. Averages of the 10 consecutive postsynaptic



potentials evoked in each treatment are shown next to the equivalent period in the graph.

FIG. 4 ATP and adenosine modulate voltage-dependent conductances of spinal neurons. **a**, ATP (10 μ M) reduced the K^+ currents evoked from a holding potential of -80 mV. Trace 1, control; 2, ATP-3; ATP-2 μ M PPADS. **b**, Adenosine reduced the Ca^{2+} currents evoked by ramps (left) or steps (right). **c**, The proposed feedback interaction between ATP and adenosine to regulate the excitability of spinal neurons (see text for further details). Computer simulations suggest that modulation of the voltage-gated currents is sufficient to cause changes in episode length. **d**, Simulation of beginning of swimming episode (ATP high, adenosine low; voltage-gated K^+ conductances reduced by 10%; Ca^{2+} conductance unchanged). **e**, Simulation of end of an episode (ATP low, adenosine high; K^+ conductances returned to resting level; Ca^{2+} conductance reduced by 10%).



to a gradual increase in the K^+ currents. Excitability would thus fall until further motor activity was no longer possible (Fig. 4c). To test the plausibility of this hypothesis, we used computer simulations of the *Xenopus* spinal circuitry that incorporate the known ion channels and synaptic connections. These simulations can reproduce many aspects of the motor pattern in the real embryo, such as cycle period and its dependence on the magnitudes of synaptic conductances¹². However, they lack the mechanisms that underlie the time-dependent changes that occur during swimming¹², and so cannot reproduce them. Nevertheless, the beginning and end of a swimming episode can be modelled according to our hypothesis (Fig. 4c) by separately reducing the K^+ and Ca^{2+} conductances by 10% (Fig. 4d, e). The simulations showed that modest reduction of the Ca^{2+} conductance is sufficient to cause early termination of swimming (Fig. 4e). Reductions of the K^+ and Ca^{2+} conductances also appropriately changed the swimming cycle period: measured over the first four cycles of simulation, the cycle period was 75 ± 1.7 ms in the control; 73 ± 1.7 ms when the K^+ currents were reduced by 10% (as at the beginning of an episode); and 79 ± 2.8 ms when the Ca^{2+} currents were reduced by 10% (as at the end of an episode). Similar increases in cycle period were found experimentally from application of adenosine to *Xenopus* embryos (beginning cycle periods, control 69 ± 1.9 ms, adenosine 74 ± 3.2 ms; $n = 8$ embryos).

Previously simulation¹² and experimental¹³ studies show that operation of the spinal circuitry in *Xenopus* is very sensitive to the balance of inward and outward currents; for example, use of K^+ channel blockers to reduce the K^+ currents by 20% or more severely disrupts the swimming motor pattern¹³. The modest effects of ATP and adenosine on the voltage-gated currents may thus represent the largest changes that can be made to modulate swimming while still preserving the correct locomotor pattern and thus efficiency of motion. □

Methods

Extracellular recordings. Stage 37/38 *Xenopus* embryos¹⁴ were prepared for extracellular ventral root recordings as previously described¹⁵. Drug access was facilitated by bilateral removal of the rostral myotomes and loosening the dorsal attachment of the remaining myotomes. The spinal cord was transected at the first post-otic myotome. The embryo was placed in a small bath (0.5 ml volume) continually superfused with saline containing (in mM): 115 NaCl, 3 KCl, 1 MgCl₂, 1 CaCl₂, 2.4 NaHCO₃, 10 HEPES, pH 7.4 at room temperature. Drugs

were applied by superfusion of the bath. Electrical stimuli to the skin were used to trigger swimming with an interval of 3 min between the beginning and end of consecutive episodes.

Patch-clamp recordings from isolated neurons. Recordings were made from neurons acutely isolated from the *Xenopus* embryo spinal cord¹⁶ that had multipolar and commissural neuron-like morphologies, and would thus have constituted a mixture of motor neurons and excitatory and inhibitory premotor interneurons¹⁷. For recording K^+ currents and drug responses, the external medium contained (in mM): 115 NaCl, 2.4 NaHCO₃, 3 KCl, 10 CaCl₂, 1 MgCl₂, 10 HEPES, pH 7.4; the pipette solution contained 100 KMeSO₃, 5 MgCl₂, 10 HEPES, 5 ATP and 2 BAPTA, pH 7.4. Ca^{2+} currents were isolated by use of an external recording medium that contained 57.5 mM TEA-Cl (substituted for an equivalent amount of NaCl) and 1 mM 4-aminopyridine in conjunction with a pipette solution that contained (in mM): 100 CsMeSO₃, 1 CaCl₂, 6 MgCl₂, 20 HEPES, 5 ATP and 10 EGTA, pH 7.4. Drugs were applied by means of a multibarrelled microprefusion system. For applications of NMDA, Mg²⁺ in the recording medium was substituted by an equivalent amount of Ca²⁺.

Intracellular recordings. Embryos were prepared as for extracellular recordings. Recordings were made from the ventral portion of the spinal cord, where most neurons are motor neurons, with potassium acetate or KCl-filled microelectrodes (150–300 M Ω). Resting input resistance was monitored by injecting hyperpolarizing current pulses of 0.2 nA. Measurements of the synaptic drive for swimming were made relative to the resting potential from 25 consecutive normal cycles. The EPSP was measured at the point of inflection on the rising phase of the spike, the IPSP from the peak of the mid-cycle IPSP, and the plateau just before the mid-cycle IPSP. Extracellular stimulation of the lateral axons was used to evoke IPSPs in motor neurons. Glycinergic IPSPs were isolated by a mixture of 20 μ M bicuculline, 100 μ M DNQX (6,7-dinitroquinoxaline-2,3-dione), 25 μ M D-2-amino-5-phosphonopentanoic acid (AP-5) and 10 μ M mecamylamine, while the glutamatergic EPSPs were isolated by a mixture of 0.5 μ M strychnine, 20 μ M bicuculline, 25 μ M AP-5 and 10 μ M mecamylamine. Embryos were continually superfused with saline containing (in mM): 115 NaCl, 3 KCl, 1 MgCl₂, 4 CaCl₂, 2.4 NaHCO₃, 10 HEPES, pH 7.4 at room temperature. For statistical comparison, a minimum of 10 consecutive evoked postsynaptic potentials were measured in each treatment.

Computer simulations. Previously described mathematical models and methods were used^{12,16}. The swimming circuit was reduced to two neurons that made self-excitatory connections (through NMDA and α -amino-5-hydroxy-5-methylisoxazole-4-propionic acid (AMPA) receptors) and reciprocal inhibitory connections (through glycine receptors). The neurons incorporated the previously described voltage-gated currents at realistic densities. The parameters to specify the simulation were those used in previous studies¹².

Received 11 June; 6 August 1996.

- McClellan, A. D. & Grillner, S. *Brain Res.* **269**, 237–250 (1983).
- Iwahara, T., Atsuta, Y., Garcia-Rill, E. & Skinner, R. D. *Somatosens. Mot. Res.* **8**, 281–287 (1991).
- Robertson, G. A., Mortin, L. I., Keifer, J. & Stein, P. S. G. *J. Neurosci.* **5**, 1517–1534 (1985).

4. Jean, A. J. *Autonom. Nerv. Syst.* **10**, 225–233 (1984).
5. Shefchyk, S. J. *Neurosci. Lett.* **99**, 175–180 (1989).
6. McKenna, K. E., Chung, S. K. & McVary, K. T. *Am. J. Physiol.* **261**, 1276–1285 (1991).
7. Wall, M. J. & Dale, N. J. *Physiol. (Lond.)* **487**, 557–572 (1995).
8. Fredholm, B. B. et al. *Pharmacol. Rev.* **46**, 143–155 (1994).
9. Lambrecht, G. et al. *Eur. J. Pharmacol.* **217**, 217–219 (1992).
10. Hoyle, C. H., Knight, G. E. & Burnstock, G. *Br. J. Pharmacol.* **99**, 617–621 (1990).
11. Gordon, E. L., Pearson, J. D. & Slakey, L. L. *J. Biol. Chem.* **261**, 15496–15504 (1986).
12. Dale, N. J. *Physiol. (Lond.)* **489**, 489–510 (1995).
13. Wall, M. J. & Dale, N. J. *Neurophysiol.* **72**, 337–348 (1994).
14. Nieuwkoop, P. D. & Faber, J. *Normal Tables of Xenopus laevis* (Daudin) (North Holland, Amsterdam, 1956).
15. Kahn, J. A. & Roberts, A. J. *Exp. Biol.* **99**, 185–196 (1982).
16. Dale, N. J. *Physiol. (Lond.)* **489**, 473–488 (1995).
17. Dale, N. *Eur. J. Neurosci.* **3**, 1025–1035.

ACKNOWLEDGEMENTS. We thank F. Kuenzi for reading earlier versions of this paper and the Royal Society and the MRC for support.

CORRESPONDENCE and requests for materials should be addressed to N.D. (e-mail: ned@st-and.ac.uk).

Role of the inositol phosphatase SHIP in negative regulation of the immune system by the receptor Fc γ RIIB

Masao Ono*, Silvia Bolland*, Paul Tempst & Jeffrey V. Ravetch*

Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

IMMUNE complexes are potent activators of inflammatory cells, triggering effector responses through the crosslinking of Fc receptors (FcRs) such as Fc ϵ RI or Fc γ RIII (ref. 1). On B cells and mast cells, immune complexes are also negative regulators of activation triggered by antigen and Fc receptors, a consequence of coligation of the B-cell antigen receptor or Fc ϵ RI, respectively, and the inhibitory receptor Fc γ RIIB. Here we show that inhibitory signalling by Fc γ RIIB does not require the SH2-domain-containing protein tyrosine phosphatase, SHP-1, in mast cells and results in the recruitment of the SH2-domain-containing inositol polyphosphate 5-phosphatase, SHIP, to the tyrosine-phosphorylated 13-amino-acid inhibitory motif of Fc γ RIIB in both B cells and mast cells. SHIP, by hydrolysing the 5-phosphate of phosphatidylinositol(3,4,5)P₃ and inositol(1,3,4,5)P₄, suggests a mechanism by which Fc γ RIIB can inhibit calcium influx and downstream responses triggered by immune receptors.

Coligation of Fc γ RIIB with Fc ϵ RI and Fc γ RIII on mast cells inhibits receptor-triggered degranulation *in vivo*² and *in vitro*^{3,4}. The SH2-domain-containing protein tyrosine phosphatase, SHP-1, has been detected associated with the activated Fc γ RIIB receptor in B cells⁵. Mice deficient in this molecule (*me/me*) have a complex immunodysregulation phenotype, including auto-antibody production and B-cell hyper-responsiveness, attributed, in part, to the loss of Fc γ RIIB inhibitory signalling⁶. To investigate whether the mechanism of negative regulation by Fc γ RIIB in mast cells is dependent upon SHP-1, bone-marrow-derived mast cells (BMMCs) from wild-type, Fc γ RII^{-/-}, *me/me* and *me^v/me^v* mice were sensitized with biotinylated IgE and triggered by crosslinking with streptavidin (Fig. 1). Increasing Fc γ RIIB crosslinking to Fc ϵ RI–IgE complexes resulted in increasing inhibition of degranulation, as measured by serotonin release, in all mast-cell preparations, with the exception of Fc γ RII knockout mice. The ability of Fc γ RIIB to inhibit IgE-triggered degranulation is indistinguishable in wild-type and mast cells derived from SHP-1-deficient mice. These results contrast with reports that

SHP-1-deficient B cells are unable to mediate Fc γ RIIB inhibition of B-cell antigen receptor (BCR) activation^{5,6}.

We next determined whether the mechanism of Fc γ RIIB inhibition in mast cells was similar to that reported for B cells^{7,8}, where receptor phosphorylation mediates the inhibitory response, resulting in reduced calcium influx. Coligation of Fc ϵ RI to Fc γ RIIB in BMMCs results in the tyrosine-phosphorylation of Fc γ RIIB (Fig. 2a), as in A20 B cells. Fc γ RIIB reconstitution in the RBL (rat basophilic leukaemic) cell line⁴ indicated that the Fc γ RIIB inhibitory activity is dependent on the same 13-amino-acid motif as for B cells, EAENTITYSLK⁸, which in turn is critically dependent on the tyrosine residue for its activity⁸. The magnitude of the calcium response following IgE crosslinking in wild-type mast cells is decreased when 2.4G2 is added, an antibody that recognizes the extracellular domains of Fc γ RII and Fc γ RIII (Fig. 2c). By contrast, in mast cells derived from Fc γ RII-deficient mice, the intensity of the IgE-triggered response increases when 2.4G2 is added, demonstrating the additive effect of Fc γ RIII signalling. Therefore, crosslinking with 2.4G2 on cells expressing both Fc γ RII and Fc γ RIII will result in a signal whose intensity is dependent on the balance between both positive and negative signalling. In contrast to B cells, where only Fc γ RIIB is expressed, stimulation with 2.4G2 alone induces calcium response in mast cells from both wild-type and Fc γ RII-deficient mice, the result of Fc γ RIII crosslinking⁹. Again, as was found for B cells, Fc γ RIIB signalling in mast cells results in the inhibition of calcium influx, while not affecting the release of intracellular stores of calcium (Fig. 2c). Stimulation of Fc ϵ RI and Fc γ RIII, either separately or simultaneously in the presence of EGTA results in a calcium response that is not detectably different in wild-type or Fc γ RII-deficient mast cells (data not shown), indicating that there is no demonstrable additive effect on the signal derived from intracellular mobilization of calcium, in contrast to the signal derived from calcium influx. Consistent with these results, receptor-triggered phosphorylation of the phospholipase PLC- γ 1 was unaffected by Fc γ RIIB coligation, again as in B cells (Fig. 2b). The mechanism of Fc γ RIIB inhibition in B cells and mast cells, for early events like receptor phosphorylation and calcium mobilization, thus appears to be similar.

As SHP-1 was not required for Fc γ RIIB-mediated inhibition in mast cells, the protein species recruited to the phosphorylated Fc γ RIIB receptor in mast cells were defined by affinity purification using tyrosine-phosphorylated and non-phosphorylated EAENTITYSLK⁸ peptides. Extracts BMMCs were incubated

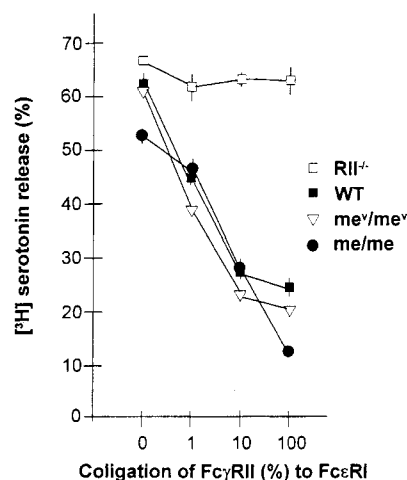


FIG. 1 Inhibitory effect of Fc γ RII coligation to Fc ϵ RI in bone-marrow-derived mast cells (BMMCs) from C57BL/6J(WT), Fc γ RII^{-/-} (RII^{-/-}), *me^v/me^v* and *me/me* mutant mice. Crosslinking of Fc ϵ RI to Fc γ RII was done as described in Methods, varying from 0 to 100% coligation. Standard errors of triplicate samples are indicated on each value.

* Present address: The Rockefeller University 1230 York Avenue, New York, New York 10021, USA.

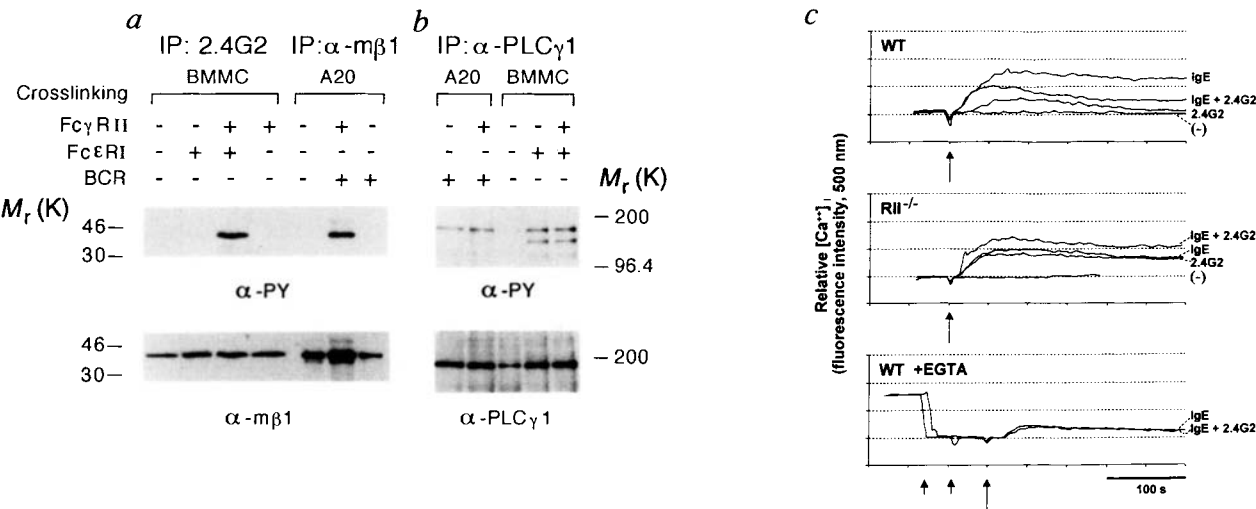


FIG. 2 Tyrosine phosphorylation (a,b) and calcium mobilization (c) in sensitized and activated mast cells on FcγRIIB coligation. Immunoprecipitates (IP) were analysed by western blotting with anti-phosphotyrosine (α-PY),

α-FcγRIIB (α-mβ1) or anti-PLCγ1 (α-PLCγ1) as indicated. Long and short arrows indicate time points of adding streptavidin for crosslinking and EGTA for depleting extracellular calcium, respectively in c.

with these peptides and revealed the presence of a prominent species at relative molecular mass 145,000 (*M_r* 145K), which associated only with the tyrosine-phosphorylated peptide (Fig. 3a, lane 1). As reported previously⁵, A20 B-cell extracts yield protein species of 160K and 65K (lane 5), whereas a murine mast-cell line, MC9, gave species of 160K and 145K (lane 3). The 65K protein is SHP-1, as confirmed by western blot analysis using an antiserum against this protein, and was absent in mast cells derived from *me/me* mice (results not shown). In contrast, the

145K protein (p145) was found associated with phosphorylated FcγRIIB peptide in mast-cell extracts derived from *me/me* mice, indicating that association of p145 is not dependent upon the presence of SHP-1. The 160K, 145K and 65K species react on western blots with an anti-pTyr antiserum (Fig. 3d). The p160 and p145 species were purified from MC9 mast cells by affinity chromatography on a tyrosine-phosphorylated FcγRIIB peptide column, transferred to nitrocellulose, and characterized by peptide mapping, mass spectrometry and protein sequence

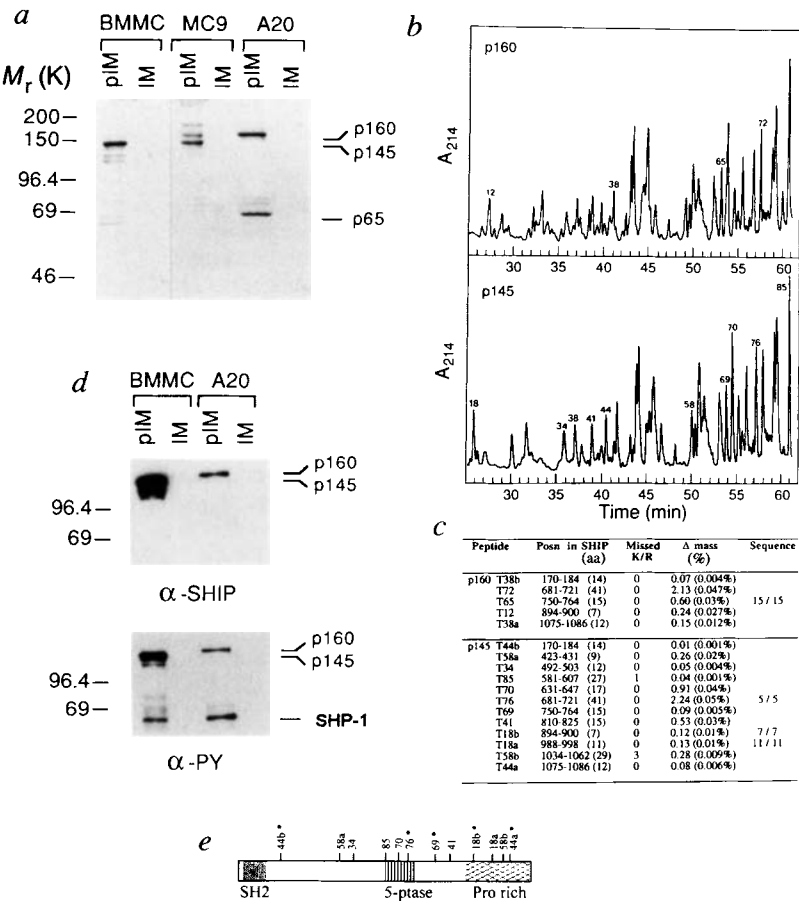


FIG. 3 Characterization of the 145K and 160K proteins associated with phosphorylated FcγRIIB-inhibitory motif. a, Lysates of BMMCs, murine mastocytoma cells (MC9) or A20 cells, adsorbed with non-phosphorylated (IM) or phosphorylated (pIM) FcγRIIB inhibitory motif peptide and silver-stained. b, Tryptic peptide maps of p160 and p145. c, Structural analysis of p160/p145 derived peptides, showing protein identity to SHIP. Missed *K/R* is the number of skipped Lysine or Arginine cleavage sites in particular peptide as listed. d, BMMCs or A20 peptide adsorbates analysed by immunoblotting with anti-SHIP or anti-phosphotyrosine (anti-PY, 4G10) antibodies. e, Representation of SHIP protein¹⁰, indicating three functional domains and the localization of peptides analysed by mass spectrometry. Numbers above the protein map correspond to p145 tryptic peptides; an asterisk indicates those that match an identical peptide in p160.

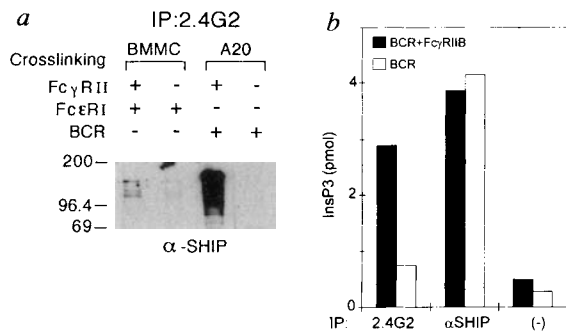


FIG. 4 Association of SHIP with FcγRII upon FcR or BCR coligation. Detection of SHIP in 2.4G2 immunoprecipitates (IP) from coligated BMMCs (4×10^7) or A20 cells (1×10^8) by *a*, immunoblotting with anti-SHIP antibodies and *b*, inositol 5-phosphatase assay.

determination (Fig. 3*b, c*). p160 and p145 had comparable peptide maps, as determined by mass analysis of isolated peptides (Fig. 3*b*). Fragment-mass fitting and sequence determination of the indicated peptides demonstrated complete identity to the SH2-containing inositol polyphosphate 5-phosphatase, SHIP (ref. 10; Fig. 3*c, e*). Identification of p160 and p145 was confirmed by western blot analysis using antiserum against the N-terminal portion of SHIP (Fig. 3*d*). SHIP exists in multiple alternatively spliced forms¹¹, which may account for the M_r difference in B cells and mast cells.

To determine whether SHIP associates with the intact FcγRIIB receptor in mast cells and B cells, we immunoprecipitated FcγRIIB after coligation to FcεRI or BCR, respectively, then did a western blot with antiserum to SHIP. As seen in Fig. 4*a* (lanes 1 and 3), SHIP associates with the FcγRIIB receptor preferentially upon coligation of the receptor to BCR or FcεRI. SHIP hydrolyses the 5-phosphate of inositol(1,3,4,5)phosphate (InsP₄) to generate inositol(1,3,4)phosphate, but is inactive against inositol(1,4,5)phosphate (InsP₃)¹⁰. SHIP activity was associated with FcγRIIB immunoprecipitates only after coligation of the receptor to BCR in A20 cells (Fig. 4*b*). Total cellular SHIP activity was not affected by BCR or FcεRI triggering (Fig. 4*b*, and data not shown). FcγRIIB thus appears to function in the recruitment of SHIP activity to the membrane, where it may act on inhibiting cellular activation.

Our data indicate that SHP-1 is not required for FcγRIIB-mediated inhibition of calcium mobilization and degranulation in BMMCs (Fig. 1). In B cells, both SHIP and SHP-1 associate with phosphorylated FcγRIIB peptide (Figs 3, 4). We propose that SHIP recruitment in B cells may serve a similar purpose to that which it does in mast cells, inhibiting early activation events like calcium influx. Previous studies on SHP-1 in B cells concluded that SHP-1-deficient *me/me* B cells did not demonstrate anti-proliferative activity on FcγRIIB engagement, suggesting a potential role for SHP-1 in modulating the late inhibitory events.⁶ Our results suggest that SHIP and SHP-1 may mediate separate responses in this inhibitory pathway, with SHIP recruitment being responsible for early activation responses dependent on calcium mobilization, like degranulation, and SHP-1 recruitment governing late responses like proliferation. Deletion and reconstitution studies will be necessary to test this hypothesis.

The inhibitory effect of FcγRIIB coligation involves the recruitment of SHIP to the membrane, where it may act on InsP₄ to inhibit its ability to stimulate the calcium channel I_{crac} (calcium release-activated channel)¹² and thereby inhibit calcium influx. SHIP might also hydrolyse phosphatidylinositol(3,4,5)phosphate (PtdInsP₃), which may function as a second messenger in mediating cellular activation events^{13,14}. As SHIP associates with Shc^{10,15}, it could also provide a mechanism for linking FcγRIIB to the Ras signalling pathways. Through its interaction with several signal-

transducing proteins, SHIP could act as a link in FcγRIIB signalling to mediate the pleiotropic inhibitory effects demonstrated by this receptor.

Methods

Mast cell isolation and FcR crosslinking. BMMCs (5×10^5 per ml), cultured with 2 ng ml^{-1} recombinant IL-3 (Genzyme), were labelled overnight with $0.4 \mu\text{M}$ 5-hydroxy[^3H]tryptamine creatinine (514 GBq per mmol; Amersham), then activated by receptor crosslinking and assayed for [^3H]serotonin release as described¹⁶. FcεRI-triggered mast-cell activation was induced by 1 h sensitization with $2.5 \mu\text{g ml}^{-1}$ biotinylated mouse IgE, followed by crosslinking with $10 \mu\text{g ml}^{-1}$ streptavidin (0% of FcγRII coligation in Fig. 1). FcγRII was coligated to FcεRI by adding biotinylated anti-FcγRII monoclonal antibody (biotin-2.4G2) at the sensitization step (100% of FcγRII coligation in Fig. 1). A mixture of biotinylated/non-biotinylated 2.4G2 (1/9 for 10% or 1/99 for 1% coligation) was used to vary the extent of FcγRII coligation.

Immunoprecipitations. Mast cells were sensitized for 1 h with anti-trinitrophenyl (TNP) monoclonal IgE ($2.5 \mu\text{g ml}^{-1}$; Sigma) followed by stimulation with $1 \mu\text{g ml}^{-1}$ TNP-conjugated ovalbumin (resulting in FcεRI crosslinking), or TNP-OVA/rabbit anti-OVA IgG immune complexes (resulting in FcγRII coligation to FcεRI). A20 cells were stimulated for 2 min with $25 \mu\text{g ml}^{-1}$ rabbit anti-IgG, F(ab')₂ fragment (resulting in BCR crosslinking), or $50 \mu\text{g ml}^{-1}$ intact antibody (resulting in FcγRII coligation to BCR). BMMCs (10^7) or A20 cells (2×10^7) were lysed in 1% NP-40 buffer⁵ and incubated for 30 min with $10 \mu\text{l}$ (50% slurry) antibody-conjugated Sepharose (2.0 mg ml^{-1} beads), or antibody plus protein A-Sepharose (Pharmacia).

Calcium measurements. BMMCs (5×10^6 in 1 ml), sensitized as described, were incubated with $3 \mu\text{M}$ Fura-2 (Molecular Probe) at 35°C for 30 min. Sensitization with IgE and 2.4G2 (a monoclonal antibody that recognizes the extracellular domains of murine FcγRII and FcγRIII) results in crosslinking FcεRI and, both FcγRII and FcγRIII, respectively. Cells were then washed, resuspended in HBSS containing 1 mM CaCl₂ and 1 mM MgCl₂, and stimulated by addition of $10 \mu\text{g ml}^{-1}$ streptavidin. Cytosolic calcium was detected as fluorescence intensity at 500 nm, excited by 340 nm. EGTA (final concentration, 5 mM) was added to quench extracellular free calcium. Intracellular calcium concentration was calculated by the formula $[\text{Ca}^{2+}] = K_d \text{Fura-2} (I - I_{\text{min}}) / (I_{\text{max}} - I)$ (ref. 17). Calcium concentrations were 196 nM (WT), 184 nM (RII^{-/-}) before crosslinking, and 437 nM (WT), 379 nM (RII^{-/-}) at the peak point after crosslinking of IgE-sensitized cells.

Protein purification. 5×10^7 cells (BMMCs) or 10^8 cells (MC9, A20) were lysed in 0.5 ml lysis buffer as described⁵, preadsorbed with protein A-Sepharose to decrease nonspecific binding, and incubated for 1 h at 4°C with IM- or pIM-conjugated Sepharose (40 mg EAENTITYSLKX peptide in $15 \mu\text{l}$ of a 50% Sepharose slurry; Biosynthesis). After extensive washing with lysis buffer, peptide adsorbates were fractionated by 10% SDS-PAGE and either silver-stained or transferred to PVDF membranes (BioRad) for immunodetection.

Protein structure determination. p145 and p160 were purified from an MC9 cell (10^9) lysate by affinity chromatography using pIM-Sepharose as described above. Proteins were eluted with 100 mM pNPP and with 100 mM glycine, pH 2.5, separated by SDS-PAGE and transferred to nitrocellulose. After *in situ* tryptic digestion, peptides were fractionated on a 1.0-mm column by RP-HPLC¹⁸, and selected peak fractions were analysed by matrix-assisted laser-desorption/ionization time-of-flight mass spectrometry using a Voyager RP Analyzer (Vestec/PerSeptive Biosystems) with TDS 520 Tektronix oscilloscope and two-peptide internal calibration¹⁹. Experimental spectrometrical mass (m/z) values for p160/p145 were used to search the NRDB (European Bioinformatics Institute, UK) using the PeptideSearch algorithm (from M. Mann). With a requirement for 11 matches out of 15, at $<0.06\%$ mass accuracy and one missed cleavage site maximum, a single protein was retrieved from the database. Mass values for p160/p145 matched, with mass differences on the average below 0.02% of those predicted from tryptic digestion of the published sequence for SHIP¹⁰. In four cases, identification was confirmed by automated Edman sequencing.

Inositol phosphatase assay. BMMCs (4×10^7) and A20 (1×10^8) cells were sensitized, crosslinked and coligated to FcγRII, followed by immunoprecipitation with 2.4G2 Sepharose as described and immunoblotting with anti-SHIP. Inositol 5-phosphatase activity was assayed in 2.4G2 Sepharose precipitates (from 5×10^8 A20 cells), or anti-SHIP + protein A-Sepharose precipitates (5×10^7 A20 cells), or protein A-Sepharose precipitates (5×10^7 cells). Reaction mixtures ($25 \mu\text{l}$) containing immunoprecipitate, 50 mM Tris-HCl, pH 7.4, 10 mM MgCl₂ and $0.5 \mu\text{M}$ *D*-myo-[2- ^3H]inositol 1,3,4,5-tetrakis-phosphate (5 Ci mmol^{-1} ; Amersham) were incubated at 37°C for 30 min and stopped by adding $500 \mu\text{l}$ cold 2 M LiCl. Samples were chromatographed on a 0.5 ml AG1-X8 formate (Bio-Rad) column equilibrated with 50 mM ammonium formate. Serial elutions were carried out starting at 0.4 M ammonium formate/ 0.1 M formic acid and ending at 1.2 M ammonium formate/ 0.1 M formic acid. Fractions containing [^3H]InsP₃ were detected according to an InsP₃ standard sample (*D*-myo-[2- ^3H]inositol 1,4,5-triphosphate; Amersham).

Received 24 May; accepted 8 August 1996.

1. Ravetch, J. V. *Cell* **78**, 553–560 (1994).
2. Takai, T. et al. *Nature* **379**, 346–349 (1996).
3. Daeron, M. et al. *J. Clin. Invest.* **95**, 577–585 (1995).
4. Daeron, M. et al. *Immunity* **3**, 635–646 (1995).
5. D'Ambrosio, D. et al. *Science* **268**, 293–297 (1995).
6. Pani, G. et al. *J. Exp. Med.* **181**, 2077–2084 (1995).
7. Amigorena, S. et al. *Science* **256**, 1808–1812 (1992).
8. Muta, T. et al. *Nature* **368**, 70–73 (1994).
9. Katz, H. R. et al. *J. Immunol.* **148**, 868–871 (1992).
10. Damen, J. E. et al. *Proc. Natl Acad. Sci. USA* **93**, 1689–1693 (1996).
11. Kavanaugh, W. M. et al. *Curr. Biol.* **6**, 438–445 (1996).
12. Luckhoff, A. & Clapham, D. E. *Nature* **355**, 356–358 (1992).
13. Toker, A. et al. *J. Biol. Chem.* **269**, 32358–32367 (1994).
14. Rameh, L. E., Chen, C. S. & Cantley, L. C. *Cell* **63**, 821–830 (1995).
15. Galandini, R. et al. *J. Exp. Med.* **183**, 179–186 (1996).
16. Daeron, M. et al. *J. Immunol.* **149**, 1365–1373 (1992).
17. Grynkiewicz, G., Poenie, M. & Tsien, R. Y. *J. Biol. Chem.* **260**, 3440–3450 (1985).
18. Erdjument-Bromage, H. et al. *Prot. Sci.* **3**, 2435–2446 (1994).
19. Elicono, C. et al. *J. Chromatogr.* **676**, 121–137 (1994).

ACKNOWLEDGEMENTS. M.O. and S.B. contributed equally to this work. We thank K. Simionovitch for *me/me* and *me'/me'* mice, G. Krystal for anti-SHIP antibodies, M. Hojo for assistance with calcium analysis, and M. Lui and H. Erdjument-Bromage for help with protein structural analysis and sequencing. These studies were supported by grants from the NIH and the DeWitt Wallace Foundation (to J.V.R.); the NSF and NCI (P.T.); and by the Charles H. Revson Foundation (M.O.).

CORRESPONDENCE and requests for materials should be addressed to J.V.R. (e-mail: ravetch@rockefeller.edu).

GTPase activity of Rab5 acts as a timer for endocytic membrane fusion

Vladimir Rybin*, Oliver Ullrich*†, Mariantonietta Rubino*, Kirill Alexandrov*, Iris Simon‡, Miguel C. Seabra§, Roger Goody‡ & Marino Zerial*

* European Molecular Biology Laboratory, Postfach 10.2209, D-69012 Heidelberg, Germany

† Institut für Biochemie, Universität Mainz, Becherweg 30, D-55128 Mainz, Germany

‡ Max-Planck Institut für Molekulare Physiologie, Rheinlanddamm 201, 44139 Dortmund, Germany

§ Department of Molecular Genetics, University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75235, USA

THE GTPase cycle is a versatile regulatory mechanism directing many cell functions¹, and Rab family members use it to regulate intracellular transport^{2–4}. Current models propose that GTP hydrolysis by Rab proteins is either required for membrane fusion or occurs afterwards to allow recycling of the protein^{1–4}. To measure the GTPase activity of Rab5 in endocytic membrane fusion⁵, we engineered a mutant that preferentially binds xanthosine 5'-triphosphate (XTP), Rab5^{D136N} and monitored the kinetics of [α -³²P]-XTP hydrolysis *in situ* during endosome fusion *in vitro*. Surprisingly, nucleotide hydrolysis occurred even in the absence of membrane fusion, indicating that membrane-bound Rab5 undergoes futile cycles of GTP (XTP) binding and hydrolysis. Nucleotide triphosphate hydrolysis by Rab5 is not conditional on membrane fusion and is reduced by its effector Rabaptin-5 (ref. 6). Our data reveal that the GTP cycle of Rab proteins differs from that of other GTPases (for example, EF-Tu) and indicate that GTP hydrolysis acts as a timer that determines the frequency of membrane docking/fusion events.

It is difficult to assess the kinetics and the requirement for GTP hydrolysis by Rab proteins *in situ* as (1) GDP–GTP exchange occurs only on binding to the membranes^{7–9}, where several different proteins bind GTP¹⁰, (2) non-hydrolysable analogues, such as GTP- γ S, do not discriminate between GTPases, and (3) mutants defective in GTP hydrolysis *in vitro* are still substrates for GTPase-activating proteins (GAPs) *in vivo*¹¹. To circumvent these

problems we engineered a mutant with modified nucleotide-binding specificity. Based on the D138N mutant of elongation factor EF-Tu¹², the corresponding mutant Rab5^{D136N} preferentially binds XTP and has a low affinity for guanine nucleotides. We calculated that the mutant protein binds XDP at least 3 orders of magnitude more tightly than GDP and shares similar nucleotide kinetic parameters with wild-type Rab5 (see Methods).

Recombinant wild-type his-Rab5 and his-Rab5^{D136N} purified from *Escherichia coli* were prenylated *in vitro* in the presence of either GDP or XDP, Rab Escort Protein (REP-1), Rab GGTase^{9,13}, and delivered to early endosome-enriched membranes. Although membrane binding of wild-type and mutant Rab5 occurred with similar kinetics and efficiency (Fig. 1a), only the membranes containing his-Rab5^{D136N} incorporated [α -³²P]-XTP in a time-dependent manner (Fig. 1b), indicating that the nucleotide specifically binds to the mutant Rab5 protein on the membrane.

We then tested the activity of Rab5^{D136N} in a cell-free assay of homotypic fusion between early endosomes that measures complex formation between internalized avidin in one endosome fraction and biotinylated horseradish peroxidase in the other. The fusion reaction is cytosol- and ATP-dependent and requires Rab5^{14,15}. As with the wild-type protein, Rab5^{D136N} was found to stimulate early endosome fusion *in vitro* (Fig. 2a). This activity strictly depended upon the presence of XTP, thus showing that nucleotide exchange on Rab5 (ref. 7) is required for membrane fusion (see also Fig. 2b), as shown previously for Ypt1p (ref. 16).

We next set out to measure the kinetics of XTP hydrolysis by membrane-bound his-Rab5^{D136N} during early endosome fusion *in vitro*. As shown in Fig. 3, in the presence of cytosol [α -³²P]-XTP was rapidly hydrolysed and after 5 min the XDP/(XDP + XTP) ratio stabilized at ~0.63, reflecting a dynamic equilibrium between XDP–XTP exchange and XTP hydrolysis. The data could be fitted with values of 0.43 min^{–1} for the XTPase reaction and 0.23 min^{–1} for the rate of XDP produced against XTP. The latter constant is 22.5-fold higher than the intrinsic dissociation rate, probably reflecting the activity of exchange factors. The XTPase rate is ~fourfold higher than the intrinsic GTPase¹⁷ and sevenfold higher

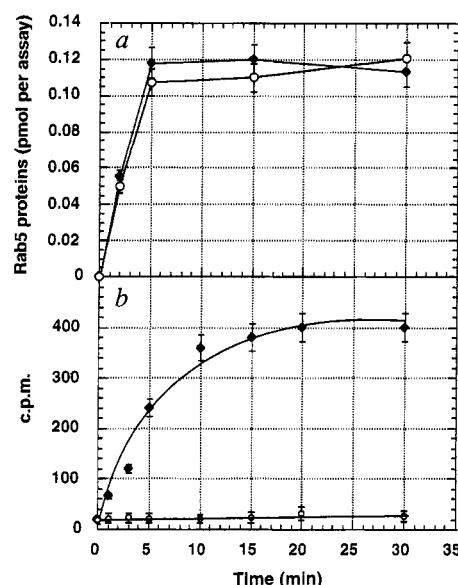


FIG. 1 Time course of binding of wild-type his-Rab5 and his-Rab5^{D136N} (a) and of [α -³²P]-XTP (b) to early endosome membranes. a, Early endosome-enriched membranes were incubated in the presence of purified REP-1:his-Rab5:GDP (○) or REP-1:his-Rab5^{D136N}:XDP (●) complexes (60 nM) for the indicated times and the amount of membrane associated Rab5 proteins (pmol) was measured by western blot analysis. In b the incorporation of [α -³²P]-XTP in early endosome membranes containing his-Rab5 (○) or his-Rab5^{D136N} (●) was determined for the indicated times.

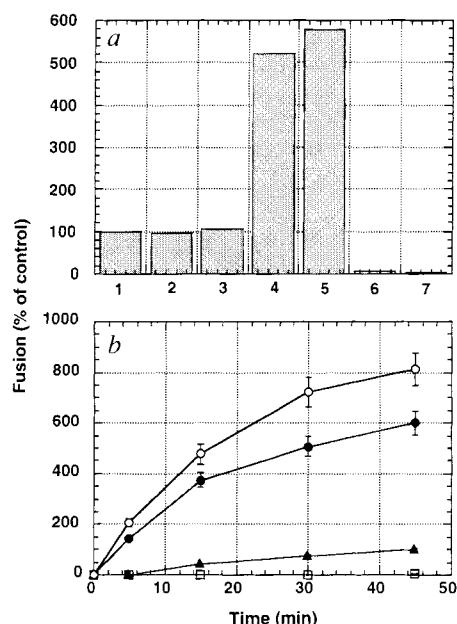


FIG. 2 REP-1:his-Rab5^{D136N} complex stimulates fusion between early endosomes *in vitro*. **a**, Fusion was conducted in the presence of cytosol alone (column 1), or cytosol supplemented with GTP (column 2), XTP (1 mM) (column 3), REP-1:his-Rab5:GDP complex (75 nM) and GTP (column 4), REP-1:his-Rab5^{D136N}:XDP complex (75 nM) and XTP (column 5), REP-1:his-Rab5^{D136N}:XDP complex and GTP (column 6), or as column 5 but in the absence of cytosol (column 7). The result of a typical experiment is shown, each value representing the mean of duplicate samples, which differed by <10%. **b**, Time course of early endosome fusion performed in the presence of REP-1:his-Rab5^{D136N}:XDP complex alone (□), or supplemented with 1 mM XTP (●) or 1 mM XTP-γS (○). The control (▲) shows fusion in the absence of complex.

than the rate of XTP hydrolysis (0.059 min⁻¹) of the purified protein, suggesting that the process is catalysed by GAPs. Nucleotide hydrolysis by Rab5^{D136N} occurred with a similar rate even in the absence of cytosol (Fig. 3), where endosome fusion is strongly inhibited (Fig. 2a), indicating that Rab5^{D136N} might undergo multiple cycles of nucleotide binding and hydrolysis. Identical kinetics occurred with membranes at one-third of the concentration, excluding a residual fusion activity of the endosomes. These data therefore demonstrate that nucleotide triphosphate hydrolysis by Rab5 is not conditional on membrane fusion.

Rabaptin-5, a recently discovered effector of Rab5, is recruited to purified early endosomes in a Rab5:GTP-dependent manner and is required for early endosome fusion⁶. We investigated whether binding of Rabaptin-5 could modify the nucleotide cycle of Rab5. Incubation in the presence of 0.5 μM recombinant Rabaptin-5, significantly decreased the rate of [α -³²P]-XTP hydrolysis (0.17 min⁻¹). Similar results were obtained with Rabaptin-5 purified from cytosol (not shown). Binding of Rabaptin-5 to Rab5 might prevent the binding of GAP, thus stabilizing the nucleotide triphosphate form.

If XTP hydrolysis occurs without membrane fusion is it required for this process? GTP-γS either inhibits or stimulates endosome fusion depending on the cytosol concentration^{15,18,19} but the factors mediating these effects are unknown. Our experimental system provides the opportunity to distinguish between the effects of the GTP analogue on Rab5 and those mediated by other GTPases. We addressed this question by using XTP-γS. When membranes were incubated in the presence of his-Rab5^{D136N}:XDP and in the absence of XTP, the early endosome fusion reaction was completely blocked (Fig. 2b), indicating that the mutant protein lacking nucleotide triphosphate is a potent inhibitor of membrane fusion. Addition of 1 mM XTP stimulated membrane fusion over the control reaction, and addition of 1 mM XTP-γS further

enhanced both the rate and the extent of endosome fusion.

These results do not rule out the possibility that endogenous Rab5 on the membrane allows a first GTPase-dependent step to proceed, followed by a second, limiting step in which Rab5:GTP is essential whereas GTP hydrolysis is not. To test this, we took advantage of the activity of Rab GDP dissociation inhibitor (RabGDI) which removes various Rab proteins from the membrane in the GDP-bound but not the GTP-bound form^{20,21}. Endosome fusion was markedly inhibited in the presence of RabGDI (Fig. 4a), which efficiently (~90%) solubilized endogenous Rab5 (Fig. 4b). However, membranes preincubated with his-Rab5^{D136N} and XTP-γS were completely resistant to RabGDI, which failed to extract the Rab5 mutant. These results demonstrate that nucleotide triphosphate hydrolysis by Rab5 is not required for membrane fusion. In addition, Rab5 seems to be a Rab protein that is both necessary and sufficient for early endosome fusion *in vitro*.

ADP-ribosylation factor (ARF) has been previously implicated both in the stimulatory and inhibitory effects of GTP-γS on endosome fusion in dilute and concentrated cytosol, respectively¹⁹. Because our data imply that GTP hydrolysis by Rab5 is not critical for endosome fusion, we tested the effect of ARF1/GTP-γS and Rab5^{D136N}/XTP-γS in the fusion reaction. Addition of 1 mM GTP-γS not only caused inhibition of early endosome fusion but also prevented stimulation by his-Rab5^{D136N} and XTP-γS (Fig. 4c). The inhibitory effect of GTP-γS at high cytosol concentration was further increased by the concomitant presence of myristoylated ARF1 (mARF1) whereas non-myristoylated ARF1 had no effect (Fig. 4d). Collectively, these data suggest that the inhibitory effect of GTP-γS on endosome fusion is not due to an inhibition of the GTPase activity of Rab5 but rather that of other GTPases, including members of the ARF family.

Our results suggesting (1) that Rab5 undergoes futile cycles of GTP (XTP) binding and hydrolysis, independently of membrane fusion and (2) that nucleotide triphosphate hydrolysis by Rab5 is not required for membrane fusion illustrate basic differences in the GTPase cycle of Rab proteins compared with other families of GTPases. For example, with EF-Tu in protein synthesis, GTP hydrolysis serves to proof-read kinetically the match between codon and anticodon and is strictly dependent on the association of the EF-Tu:aminoacyl-tRNA with the ribosome^{1,22}. The GTPase

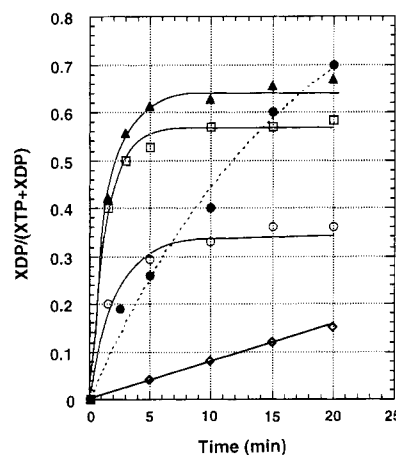
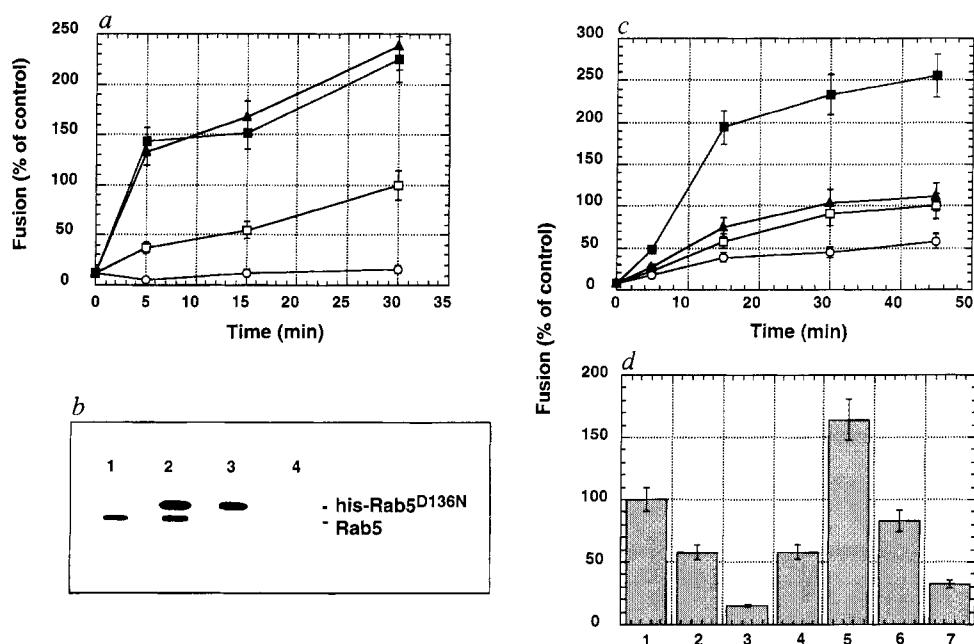


FIG. 3 Hydrolysis of [α -³²P]-XTP by Rab5^{D136N}. The purified his-Rab5^{D136N} protein was preloaded with [α -³²P]-XTP and incubated either at 4 °C (◇) or at 37 °C (●) for various periods. To measure nucleotide hydrolysis by membrane-bound his-Rab5, early endosome membranes were incubated in the presence of 60 nM REP-1:his-Rab5^{D136N} complex for 15 min at 37 °C, loaded with [α -³²P]-XTP, and incubated for various times in the absence (□), or presence of cytosol (▲) or recombinant Rabaptin-5 (○). Similar results were obtained when the membranes were incubated simultaneously with REP-1:his-Rab5^{D136N} complex and [α -³²P]-XTP (not shown).

FIG. 4 a, Rab5^{D136N}:XTP-γS rescues the RabGDI-mediated inhibition of early endosome fusion. Fusion reactions preincubated in the presence (▲, ■) or absence (□, ○) of REP-1:his-Rab5^{D136N} and XTP-γS before incubation in the presence (○, ▲) or absence (■, □) of RabGDI. b, The endosome membranes incubated as in a were analysed for the presence of endogenous Rab5 and his-Rab5^{D136N} as described in Fig. 1. Lane 1, control endosomes; lane 2, endosomes incubated with REP-1:his-Rab5^{D136N} and XTP-γS; lane 3, endosomes incubated with REP-1:his-Rab5^{D136N}, XTP-γS and RabGDI; lane 4, endosomes incubated with RabGDI alone. The presence of the his-tag causes a shift in electrophoretic mobility of the exogenous protein compared with endogenous Rab5. c, GTP-γS prevents the stimulation of early endosome fusion induced by his-Rab5^{D136N}. Fusion was conducted under normal conditions (□) or in the presence of GTP-γS (○), REP-1:his-Rab5^{D136N} complex and XTP-γS (■) or REP-1:his-Rab5^{D136N} complex, XTP-γS and GTP-γS (▲). d, mARF1 and GTP-γS counteract the stimulatory effect of his-Rab5^{D136N}:XTP-γS on early endosome fusion. Standard fusion reactions were incubated in the absence (column 1) or



presence of GTP-γS (column 2), mARF1 (column 3), ARF1 (column 4); or REP-1:his-Rab5^{D136N}:XDP complex and XTP-γS, either alone (column 5), or with GTP-γS (column 6) or GTP-γS and mARF1 (column 7).

activity of dynamin is thought to induce a mechanochemical conformational change driving clathrin-coated vesicle budding^{23,24}. Our data strongly support another model for the GTPase cycle of Rab5 in which GTP hydrolysis is not conditional on membrane fusion but rather serves to sustain a dynamic equilibrium between the GTP and GDP-bound form. This mechanism shares similarities with the GTP cycle of Ras²⁵ and of heterotrimeric G-protein α subunits that use GTP hydrolysis as a timer to amplify transmembrane signals²⁶. In this case the equilibrium between the GTP- and GDP-bound forms is regulated by receptor activation. For Rab5 we propose that the continuous alternating of GDP-Rab5 exchange and GTP hydrolysis regulates the recruitment of Rabaptin-5 on the membrane⁶. Binding of Rabaptin-5 itself retards GTP hydrolysis, presumably by blocking the activation by GAPs. Thus, once recruited on the membrane, Rabaptin-5 exerts a positive feedback effect on the endosome docking/fusion machinery remaining bound to Rab5 long enough to run the membrane fusion process to completion. The intrinsic GTPase activity would ensure that Rabaptin-5 is eventually released into the cytosol. The GTPase-deficient Rab5 mutant potentially promotes endosome fusion despite being sensitive to GAPs¹¹ probably because it delays the release of Rabaptin-5 into the cytosol.

Such dynamic GDP-GTP cycling by Rab5 provides several advantages to the membrane transport machinery. First, it is regulatable by modulating the activity of GEF and GAP, for example, by phosphatidylinositol 3-kinase²⁷. Second, GTP hydrolysis ensures that, if mis-sorted, Rab5 does not last too long in the active form leading to organelle mixing. Last, it provides temporal regulation. The estimated half-time for GTP hydrolysis by membrane-bound Rab5 is 1.5 min, a value closely matching the $t_{1/2}$ of transport from the surface into early endosomes (~ 1 min)²⁸. In this respect it is interesting that Rab5, which has one of the highest basal GTPase rates for members of the Rab family, regulates one of the fastest intracellular transport processes. The requirement for such a dynamic GTPase cycle becomes apparent if one considers the dual role of Rab5 in the fusion of clathrin-coated vesicles with early endosomes and the homotypic fusion between early endosomes^{5,15}. Downregulation of Rab5 might be necessary to maintain the correct size and distribution of the endosomal compartment and prevent the formation of

large vacuoles¹¹. On the basis of these data we propose that the timing of GTP hydrolysis, largely built into the structure of Rab GTPases, is responsible for regulating the kinetics of the membrane docking and fusion process. □

Methods

Preparation, characterization and membrane delivery of Rab5 proteins.

His-Rab5^{D136N} was purified from *E. coli* as described for his-Rab5 (ref. 9). Dissociation rates of guanosine, xanthosine and methylanthraniloyl (mant) nucleotide derivatives were determined by fluorescence measurements¹⁷. The XDP dissociation (off) rate from wild-type Rab5 is $6 \times 10^{-3} \text{ s}^{-1}$ at 25 °C. The dissociation rate constant for mantXDP from his-Rab5^{D136N} was estimated as $1.7 \times 10^{-4} \text{ s}^{-1}$ at 37 °C, and the corresponding rate of GDP from wild type as $1 \times 10^{-4} \text{ s}^{-1}$. The GDP off rate from the mutant was not measured because it was not possible to displace XDP by excess of GDP. In the wild-type protein the off rate of XDP is 1,000-fold greater than that of GDP. Thus the mutation slows down the off rate of XDP by a factor of $\sim 1,000$, and presumably speeds up the GDP off rate by a similar factor. In comparison with studies with the analogous Ras^{D119N} mutant there is probably no effect on the on rate, and we can conclude that XDP binds at least 3 orders of magnitude more tightly than GDP to the mutant and with similar affinity to that of GDP to wild-type Rab5. For membrane delivery assays, purified his-Rab5^{D136N} was prenylated *in vitro* in the presence of Rab GGTase in a reaction that required XDP^{6,13}. Aliquots (10 μl) of early endosome membranes¹⁵ in buffer B (12.5 mM HEPES, pH 7.2, 75 mM K acetate, 1 mM Mg acetate, 1 mM dithiothreitol) with 30 mM ATP (to prevent unspecific hydrolysis of XTP owing to cytosolic phosphatases), 1 mM Na phosphate and Na pyrophosphate, pH 7.2, were incubated at 37 °C with 60 nM REP-1:his-Rab5 or REP-1:his-Rab5^{D136N} complex⁹ for the indicated periods. Samples were then diluted 40-fold by ice-cold buffer B containing 20 mM MgCl₂, membranes were pelleted (200,000g) for 10 min at 2 °C, washed twice and analysed by immunoblotting with anti-Rab5 monoclonal antibody.

Measurement of [α -³²P]-XTP incorporation and hydrolysis. Pre-loading of his-Rab5^{D136N} with [α -³²P]-XTP and TLC separation and quantification were as described for his-Rab5:GTP¹¹. XTP hydrolysis was calculated as $R_t = (A_t - A_0) / (1 - A_0)$, where A_0 is the ratio $[XDP] / ([XDP] + [XTP])$ at time 0 and A_t is the same ratio at time t . To calculate the rate constant of XTP hydrolysis by his-Rab5^{D136N}, we took into consideration that rate constants of dissociation of GTP and GDP from Rab5 are similar and very small²⁷. In this case it is possible to use the following equation:

$$R_t = 1 - e^{-k_{\text{hyd}} t} \quad (1)$$

where k_{hyd} is hydrolysis rate constant. For XTP incorporation and hydrolysis by membrane-bound his-Rab5^{D136N}, early endosome-enriched membranes pre-incubated with REP-1:his-Rab5 or REP-1:his-Rab5^{D136N} complex (60 nM) for 15 min at 37 °C were incubated with 0.5 μM [α -³²P]-XTP (3,000 Ci mmol⁻¹,

Transcription-linked acetylation by Gcn5p of histones H3 and H4 at specific lysines

Min-Hao Kuo*, James E. Brownell*, Richard E. Sobel*†, Tamara A. Ranalli*, Richard G. Cook‡, Diane G. Edmondson§, Sharon Y. Roth§ & C. David Allis*

* Department of Biology, University of Rochester, Rochester, New York 14627, USA

‡ Department of Microbiology and Immunology, Baylor College of Medicine, Houston, Texas 77030, USA

§ Department of Biochemistry and Molecular Biology, M. D. Anderson Cancer Center, Houston, Texas 77030, USA

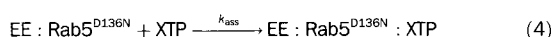
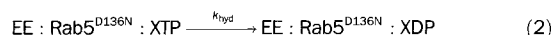
THE yeast transcriptional adaptor¹⁻³, Gcn5p, is a catalytic subunit of a nuclear (type A) histone acetyltransferase linking histone acetylation to gene activation⁴⁻⁶. Here we report that Gcn5p acetylates histones H3 and H4 non-randomly at specific lysines in the amino-terminal domains. Lysine 14 of H3 and lysines 8 and 16 of H4 are highly preferred acetylation sites for Gcn5p. We also demonstrate that lysine 9 is the preferred position of acetylation in newly synthesized yeast H3 *in vivo*. This finding, along with the fact that lysines 5 and 12 in H4 are predominant acetylation sites during chromatin assembly of many organisms⁷⁻¹¹, indicates that Gcn5p acetylates a distinct set of lysines that do not overlap with those sites characteristically used by type B histone acetyltransferases for histone deposition and chromatin assembly.

To understand the substrate specificity of recombinant Gcn5p (rGcn5p) under our standard *in vitro* assay conditions, we characterized the acetylation pattern of rGcn5p with specific histone substrates. When yeast core histones and ³H-acetyl coenzyme A are incubated in the presence of rGcn5p, significant ³H-acetate counts are incorporated into trichloroacetic acid (TCA)-insoluble product in a histone- and rGcn5p-dependent fashion (Fig. 1a). In agreement with our earlier results, no acetylation is detected when bovine serum albumin (BSA) is substituted in place of histones⁴.

To determine the substrate specificity, histones from various organisms were acetylated *in vitro* with rGcn5p and analysed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and fluorography. Under these conditions, H3 from all organisms tested is acetylated strongly (Fig. 1b). In contrast, H4 is weakly acetylated; H2A and H2B are seldom acetylated above 'minus enzyme' controls. The relatively weak labelling of H4 is, at least in part, due to the use of a roughly equal molar mixture of core histones in the above reactions. When reversed-phase high-performance liquid chromatography (HPLC)-purified H4 from *Tetrahymena* (Fig. 1c) or yeast (not shown) is incubated with rGcn5p, a much greater acetylation efficiency is typically observed.

To identify rigorously the lysine(s) acetylated by rGcn5p, we subjected *in vitro* rGcn5p-labelled yeast H3 to microsequence analyses followed by direct determination of radioactivity at each position in the H3 sequence. In this analysis, ³H-label is in the acetyl moiety itself, easily defining the acetylated residues. The results show clearly that Lys 14 is the major, if not exclusive, site of rGcn5p acetylation in yeast H3 *in vitro* (Fig. 2a). Two lines of evidence argue that the nearly exclusive use of Lys 14 in yeast H3 is not due to differences in site availability of the lysines in H3 itself. First, most of the yeast H3 is underacetylated (about 80% of total H3), as demonstrated by high-resolution triton-acid-urea gel electrophoresis and by the mass ratio of acetyl-lysine to lysine determined during the sequencing of bulk yeast H3 shown in

PBK9601, Amersham) for the indicated times. When added, cytosol was at a final concentration of 3.6 mg ml⁻¹ and recombinant Rabaptin-5 was 0.5 μM. All other steps were as above except that, after pelleting the membranes, buffer C (0.2% SDS, 2 mM EDTA, 10 mM XTP and 10 mM XDP, pH 7.2) and buffer D (as buffer C but with 10 mM GTP and 10 mM GDP in place of XTP and XDP) were added for his-Rab5^{D136N} and his-Rab5, respectively. Samples were heated for 2 min at 70 °C and membrane-associated radioactivity was measured by scintillation counting. For determination of the hydrolysis rate constant for membrane-bound his-Rab5^{D136N}, experiments were done as described above, except that thin-layer chromatography followed counting the radioactivity. Considering that nucleotide hydrolysis and exchange reactions take place in the same system, we have used the following model:



In our experiment the concentration of XTP was 0.5 μM, so that the effective rate of XTP binding (given by [XTP] × k_{ass}) is likely to be much higher than the rate of XDP dissociation (k_{diss}) as expected from the kinetic scheme given above, before settling down to a steady-state rate commensurate with the constants derived from the data in Fig. 3, both with and without Rabaptin-5.

$$R_t = (k_{hyd} / (k_{hyd} + k_{diss})) (1 - e^{-(k_{hyd} + k_{diss})t}) \quad (5)$$

The kinetics of XDP appearance in the cytosol (not shown) indicated a pronounced lag phase, as expected from the kinetic scheme given above, before settling down to a steady-state rate commensurate with the constants derived from the data in Fig. 3, both with and without Rabaptin-5.

Early endosome fusion assays. Fusion of bHRP- and avidin-labelled early endosome fractions was performed as described¹⁵ in the presence or absence of bovine brain cytosol (3.6 mg ml⁻¹, except in Fig. 4 at 3 mg ml⁻¹), 1 mM GTP and ATP-regenerating system. Fusion was quantified by immunoprecipitation of the avidin-bHRP complexes, with the Immuno-Pure detection liquid (Pierce, Rockford, IL, USA). XTP-γS was synthesized as described²⁹. For RabGDI inhibition experiments, standard fusion reactions were preincubated for 5 min at 37 °C in the presence or absence of 75 nM REP-1:his-Rab5^{D136N} complex and 1 mM XTP-γS without ATP regeneration system to block fusion. ATP regeneration system was then added and incubation was continued at 37 °C in the presence or absence of 0.5 μM RabGDI purified from bovine brain cytosol³⁰. Recombinant myristoylated mARF-1 and non-myristoylated ARF-1 were added at a concentration of 10 μM and GTP-γS at 1 mM.

Received 26 June; accepted 23 July 1996.

- Bourne, H. R., Sanders, D. A. & McCormick, F. *Nature* **346**, 125–132 (1990).
- Zerial, M. & Stenmark, H. *Curr. Opin. Cell Biol.* **5**, 613–620 (1993).
- Novick, P. & Brennwald, P. *Cell* **75**, 597–601 (1993).
- Pfeffer, S. R. *Curr. Opin. Cell Biol.* **6**, 522–526 (1994).
- Bucci, C. *et al.* *Cell* **70**, 715–728 (1992).
- Stenmark, H., Vitale, G., Ullrich, O. & Zerial, M. *Cell* **83**, 423–432 (1995).
- Ullrich, O., Horuchi, H., Bucci, C. & Zerial, M. *Nature* **368**, 157–160 (1994).
- Soldati, T., Shapiro, A. D., Dirac Svejstrup, A. B. & Pfeffer, S. R. *Nature* **369**, 76–78 (1994).
- Alexandrov, K., Horuchi, H., Steele-Mortimer, O., Seabra, M. & Zerial, M. *EMBO J.* **13**, 5262–5273 (1994).
- Huber, L. A. *et al.* *Proc. Natl Acad. Sci. USA* **91**, 7874–7878 (1994).
- Stenmark, H. *et al.* *EMBO J.* **13**, 1287–1296 (1994).
- Weijland, A. & Parmeggiani, A. *Science* **259**, 1311–1314 (1993).
- Seabra, M. C. *J. Biol. Chem.* **271**, 14398–14404 (1996).
- Diaz, R., Mayorga, L. & Stahl, P. *J. Biol. Chem.* **263**, 6093–6100 (1988).
- Govvel, J. P., Chavrier, P., Zerial, M. & Gruenberg, J. *Cell* **64**, 915–925 (1991).
- Jones, S., Litt, R. J., Richardson, C. J. & Segev, N. *J. Cell Biol.* **130**, 1051–1061 (1995).
- Simon, I., Zerial, M. & Goody, R. *J. Biol. Chem.* (in the press).
- Mayorga, L. S., Diaz, R. & Stahl, P. *D. Science* **244**, 1475–1477 (1989).
- Lenhard, J. M., Kahn, R. A. & Stahl, P. *D. J. Biol. Chem.* **267**, 13047–13052 (1992).
- Soldati, T., Riederer, M. A. & Pfeffer, S. R. *Mol. Biol. Cell* **4**, 425–434 (1993).
- Ullrich, O. *et al.* *J. Biol. Chem.* **268**, 18143–18150 (1993).
- Sprinzel, M. *Trends Biochem. Sci.* **19**, 245–250 (1994).
- Takei, K., McPherson, P. S., Schmid, S. L. & De Camilli, P. *Nature* **374**, 186–190 (1995).
- Hinshaw, J. E. & Schmid, S. L. *Nature* **374**, 190–192 (1995).
- Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. *Nature* **346**, 719–723 (1990).
- Gilman, A. G. *Cell* **36**, 577–579 (1984).
- Li, G. *et al.* *Proc. Natl Acad. Sci. USA* **92**, 10207–10211 (1995).
- Griffiths, G., Back, R. & Marsh, M. *J. Cell Biol.* **109**, 2703–2720 (1989).
- Goody, R. S., Eckstein, F. & Schirmer, H. *Biochim. Biophys. Acta* **276**, 155–161 (1972).
- Sasaki, T. *et al.* *J. Biol. Chem.* **265**, 2333–2337 (1990).

ACKNOWLEDGEMENTS. We thank A. Giner for technical assistance, P. Metcalf for valuable suggestions, H. Horuchi for providing purified RabGDI, J. Lanox and B. Hoflack for their gift of ARF1 proteins, and B. Hoflack, C. Murphy, T. Nilsson, K. Simons and P. Woodman for critical reading of this manuscript. O.U. was supported by a grant from the Deutsche Forschungsgemeinschaft.

CORRESPONDENCE and requests for materials should be addressed to M.Z. (e-mail: Zerial@EMBL-heidelberg.DE).

† Present address: Zoology Department, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada.

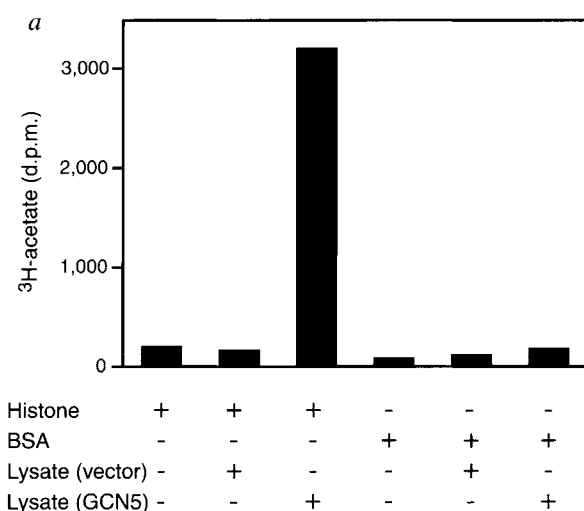
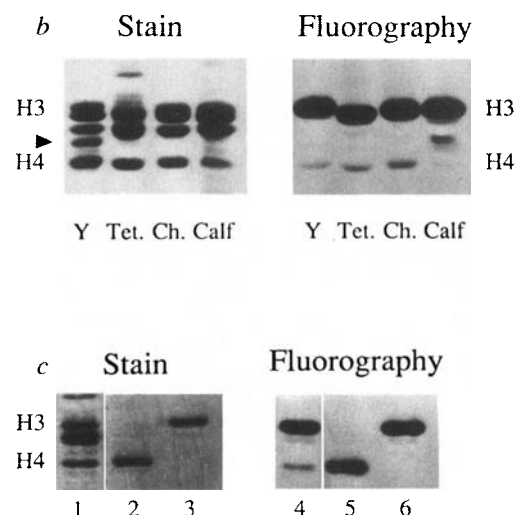


FIG. 1 Recombinant yeast Gcn5p acetylates H3 and H4 selectively. *a*, Bacterial extracts with or without rGcn5p were incubated with yeast core histones or BSA plus ³H-acetyl coenzyme A. Incorporation of ³H-label was monitored by TCA precipitation and scintillation counting. *b*, Core histones purified from various organisms were tested for their acetylation by yeast rGcn5p. Y, Yeast; Tet., *Tetrahymena thermophila* macronuclei; Ch., chicken erythrocyte; Calf, calf thymus. The positions of H3 and H4 are indicated. Fluorography was done with the same gel and the autoradiogram is shown in



the right panel. The proteolytic product of yeast H3 lacking the first 20–22 amino acids (our unpublished protein sequencing data) is marked with a closed triangle on the left panel; as expected, this species is unlabelled by rGcn5p. *c*, Reversed-phase HPLC-purified *Tetrahymena* macronuclear H3 and H4 were separately incubated with rGcn5p and the efficiency of acetylation was monitored by fluorography. Lanes 1 and 4, pre-HPLC materials; lanes 2 and 5, H4; lanes 3 and 6, H3.

Fig. 2*a* (data not shown). Second, synthetic peptides corresponding to the N-terminal domain of yeast H3 exhibit a similar rGcn5p acetylation pattern, indicating a high preference of Lys 14 even when other lysines are available (unpublished observations).

To map the site(s) of acetylation in yeast H4 by rGcn5p, we took advantage of methods recently developed to deblock this histone chemically⁸ before microsequencing. As shown in Fig. 2*b*, most of the ³H-acetate label is recovered at the position of Lys 8. Little, if any, acetylation is observed in any of the other potential acetylation sites (Lys 5, Lys 12 and Lys 16). Thus, in agreement with the results obtained with yeast H3 (Fig. 2*a*), rGcn5p is highly selective towards specific lysines in the N-terminal domains of H3 and H4.

Most of the steady-state (bulk) yeast H4 has been reported to be acetylated at Lys 16 (ref. 12), raising the possibility that site occupancy from pre-existing acetylation might prevent acetylation by rGcn5p at this site. To test this directly, deacetylated *Tetrahymena* H4 was used as an *in vitro* substrate for rGcn5p and subjected to microsequencing. As shown in Fig. 2*c*, *Tetrahymena* H4 is acetylated by rGcn5p at Lys 7 and Lys 15 (equivalent to Lys 8 and Lys 16 in other organisms) with most acetylation occurring at Lys 15. Little, if any, acetylation is detected at other potential acetylation sites (Lys 4 or Lys 11). From these data, we suggest that Lys 8 and Lys 16 are highly preferred acetylation sites in H4 for yeast Gcn5p.

The above data indicate a mutually exclusive use of acetylation sites for rGcn5p-mediated acetylation of H4 *in vitro* (Lys 7/8 and Lys 15/16; Fig. 2) as compared to deposition-related acetylation of H4 *in vivo* (Lys 4/5 and Lys 11/12 (refs 7, 13)). Because H3 is a highly preferred substrate for Gcn5p (Fig. 1*b*), we sought to determine the site(s) of deposition-related H3 acetylation in yeast to see if these sites also differed from those recognized by rGcn5p.

Exponentially growing yeast cells were pulse-labelled with ³H-lysine and subjected to histone extraction. H3, purified by reversed-phase HPLC, was then sequenced and analysed using a strategy developed to determine deposition-related acetylation sites¹³. Table 1 shows the comparison of ³H counts, derived from each sequencing cycle of lysine residues in the H3 N-terminal

domain (Lys 4, 9, 14, 18, 23 and 27), that correspond to lysine and acetyl-lysine eluting from HPLC. Among the known lysines, Lys 9 showed the highest percentage of acetyl-lysine (56%), followed by Lys 27 (36%). These results are consistent with the mobility of newly synthesized H3 in long acid-urea gels, which demonstrated a preponderance of new H3 migrating at the position of mono-acetylated H3 (R.E.S. and C.D.A., unpublished observation). Our results suggest that Lys 9 is a predominant deposition-related acetylation site in yeast H3 and, as with H4, the deposition-related acetylation site in H3 does not overlap with the highly preferred Lys 14 acetylating site used by rGcn5p *in vitro*.

Our data demonstrate that yeast Gcn5p is highly selective for Lys 14 of H3 and Lys 8 and Lys 16 of H4 when acetylating free histone substrates under our *in vitro* assay conditions. Importantly, this Gcn5p-mediated acetylation site pattern is completely non-overlapping with the previously described deposition-related site pattern (Lys 9 of H3 (Table 1), Lys 5/12 of H4 (refs 7–11)), suggesting that histone acetyltransferases of the A and B class

TABLE 1 Newly synthesized H3 in yeast is preferentially acetylated at Lys 9 during histone deposition and chromatin assembly

Lysine positions	Acetyl-lysine ³ H (c.p.m.)	Lysine ³ H (c.p.m.)	Percentage acetyl-lysine
4	0	574	0
9	644	513	56
14	153	637	20
18	33	656	5
23	139	710	16
27	328	568	37

³H-lysine pulsed-labelled yeast H3 was purified by reversed-phase HPLC and subjected to microsequencing. At the known position of lysines in the N-terminal domain of H3 (lysines at 4, 9, 14, 18, 23 and 27), fractions from the sequencer HPLC corresponding to acetyl-lysine and unmodified lysine were analysed as described in ref. 13. The c.p.m. corresponding to acetyl-lysine or unmodified lysine at each position is shown.

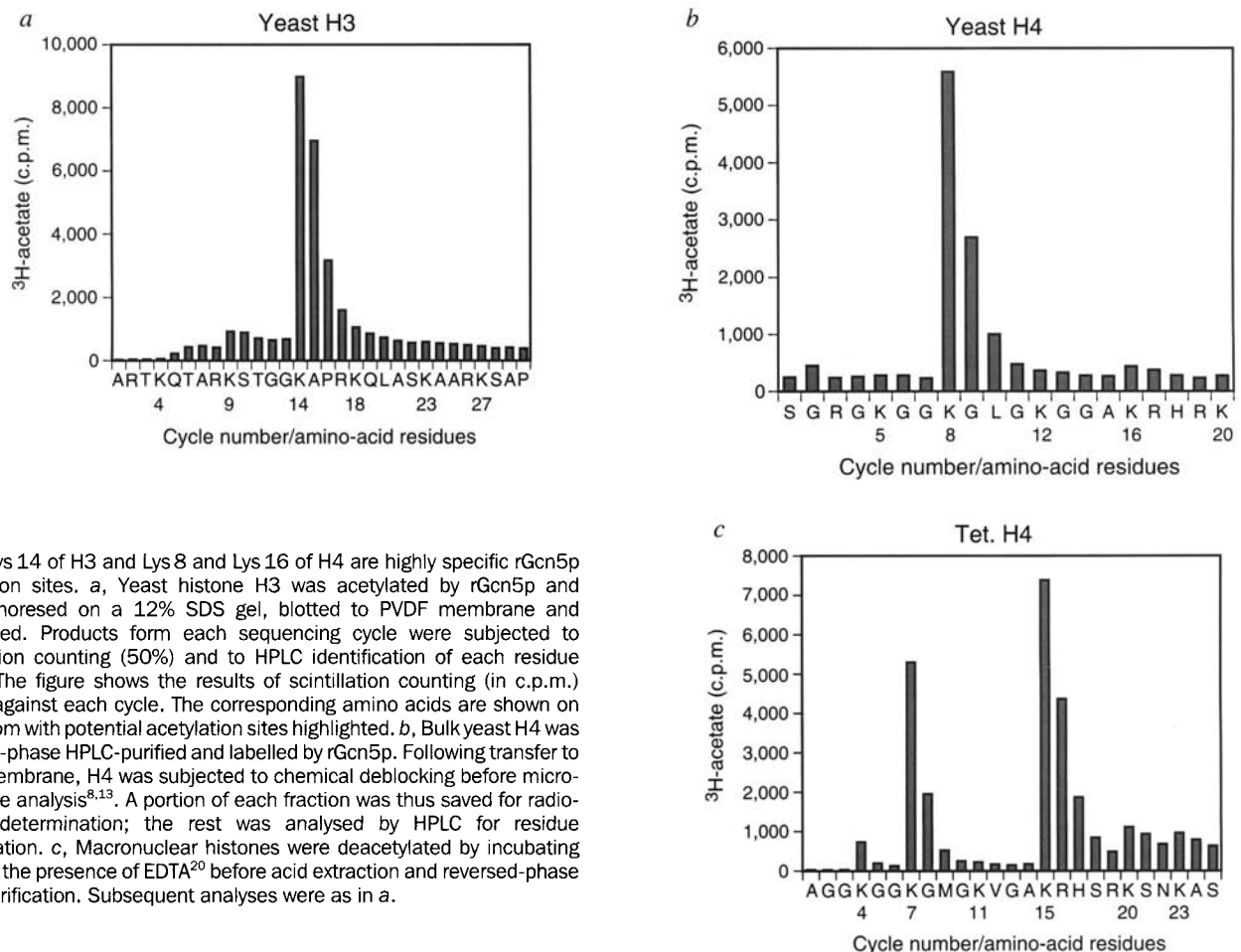


FIG. 2 Lys 14 of H3 and Lys 8 and Lys 16 of H4 are highly specific rGcn5p acetylation sites. **a**, Yeast histone H3 was acetylated by rGcn5p and electrophoresed on a 12% SDS gel, blotted to PVDF membrane and sequenced. Products from each sequencing cycle were subjected to scintillation counting (50%) and to HPLC identification of each residue (50%). The figure shows the results of scintillation counting (in c.p.m.) plotted against each cycle. The corresponding amino acids are shown on the bottom with potential acetylation sites highlighted. **b**, Bulk yeast H4 was reversed-phase HPLC-purified and labelled by rGcn5p. Following transfer to PVDF membrane, H4 was subjected to chemical deblocking before microsequence analysis^{8,13}. A portion of each fraction was thus saved for radioactivity determination; the rest was analysed by HPLC for residue identification. **c**, Macronuclear histones were deacetylated by incubating nuclei in the presence of EDTA²⁰ before acid extraction and reversed-phase HPLC purification. Subsequent analyses were as in **a**.

each catalyse acetylation at distinct sites, presumably for specific functions (transcriptional regulation versus chromatin assembly, respectively). The acetylation site pattern that we have observed with Gcn5p and free histones is completely reproduced with synthetic peptides corresponding to N-terminal histone domains. It seems likely then that much specificity for site selection used by Gcn5p resides in the N-terminal histone domain itself. Consistent with this hypothesis, careful inspection of preferred Gcn5p acetylation sites reveals a potential Gcn5p acetylation consensus sequence (K-X-X-G/A-K*-X-not G-X-K/R, where K* is the acetyl acceptor). Gcn5p is part of a transcriptional adaptor multi-subunit complex (refs 14–6 and K. J. Pollard and C. L. Peterson, manuscript in preparation), and the Ada proteins in this complex may be necessary for association with chromatin. We (J.E.B. *et al.*, manuscript in preparation) and others¹⁷ have consistently found that rGcn5p is unable to acetylate nucleosomal substrates as a purified catalytic subunit. The site specificity of this enzyme complex is likely to provide yet another level of transcriptional regulation *in vivo*. □

Methods

Strains, DNA, protein and peptides. Subcloning, expression and purification of recombinant yeast Gcn5p from *Escherichia coli* are as described⁴. Yeast strains and core histone isolation have been described in detail previously¹⁸. *Tetrahymena* histones were prepared following published procedures¹⁹; where indicated, deacetylated histones were prepared by incubating macronuclei in an EDTA-containing buffer as described²⁰. Chicken histones were a gift from C. Mizzen, and calf thymus histones were purchased from Sigma (St Louis, MO).

In vivo labelling of newly synthesized yeast histones. Methods for *in vivo* pulse labelling and purification of newly synthesized yeast histones will be published in detail elsewhere (R.S. and C.D.A., manuscript in preparation).

Briefly, yeast cells collected from mid-log phase were starved for lysine and were spheroplasted before pulse labelling with ³H-lysine for 10 min. Nuclei were isolated followed by immediate acid extraction of the core histones²¹.

In vitro acetylation. A typical 50-μl *in vitro* acetylation reaction consisted of 50 mM Tris-HCl, pH 8.0, 10% glycerol, 1 mM DTT, 10 mM *n*-butyrate, 0.2 μCi ³H-acetyl coenzyme A (5–6 Ci mmol⁻¹), 25 μg of core histones, and 2 μl of Ni-purified recombinant His-tagged Gcn5p. The reaction was incubated at 30 °C for 20 min. For incorporation counting, a portion of the reaction was subject to P81 filter method^{22,23} and resolved on a 12% SDS-PAGE. Alternatively, 20% TCA precipitation of histones were performed for the scintillation counting. Histones and acetylated histone products were visualized by Coomassie blue staining and fluorography, respectively.

Protein microsequencing. Microsequence analyses of the N-terminal tails of yeast H3 and H4 were as described^{8,13}. For H3 sequencing, core histones were labelled *in vitro* with rGcn5p and ³H-acetyl coenzyme A followed by electrophoresis on a 12% SDS polyacrylamide gel. In the case of H4, *Tetrahymena* or yeast histones were first subjected to reversed-phase HPLC. *Tetrahymena* histones were resolved⁸; yeast core histones were chromatographed on a Vydac reversed-phase C-8 column (Separations Group, Hesperia, CA) and eluted with a linear ascending concentration of acetonitrile (27% to 54% in 180 min). Under these gradient conditions, H4 coelutes with H2A. The H4/H2A fraction was lyophilized and redissolved in water followed by *in vitro* acetylation by rGcn5p. All subsequent processing steps were essentially identical to those for H3 microsequencing. In the case where H3 was pulse-labelled *in vivo* with ³H-lysine (see above), microsequencing, separation of phenylthiohydantoin derivatives of acetyl-lysine and lysine and subsequent analyses were as described¹³.

Received 17 June; accepted 26 July 1996.

- Georgakopoulos, T. & Thireos, G. *EMBO J.* **11**, 4145–4152 (1992).
- Berger, S. L., Cress, W. D., Cress, A., Triezenberg, S. J. & Guarente, L. *Cell* **61**, 1199–1208 (1990).
- Berger, S. L. *et al.* *Cell* **70**, 251–265 (1992).
- Brownell, J. E. *et al.* *Cell* **84**, 843–851 (1996).
- Wolffe, A. P. & Pruss, D. *Cell* **84**, 817–819 (1996).

6. Brownell, J. E. & Allis, C. D. *Curr. Opin. Genet. Dev.* **6**, 176–184 (1996).
7. Chicoine, L. G., Schulman, I. G., Richman, R., Cook, R. G. & Allis, C. D. *J. Biol. Chem.* **261**, 1071–1076 (1986).
8. Sobel, R. E., Cook, R. G. & Allis, C. D. *J. Biol. Chem.* **269**, 18576–18582 (1994).
9. Munks, R. J., Moore, J., O'Neill, L. P. & Turner, B. M. *FEBS Lett.* **284**, 245–248 (1991).
10. Turner, B. M., O'Neill, L. P. & Allan, I. M. *FEBS Lett.* **253**, 141–145 (1989).
11. Thorne, A. W., Kmiecik, D., Mitchelson, K., Sautiere, P. & Crane-Robinson, C. *Eur. J. Biochem.* **193**, 701–713 (1990).
12. Clarke, D. J., O'Neill, L. P. & Turner, B. M. *Biochem. J.* **294**, 557–561 (1993).
13. Sobel, R. E., Cook, R. G., Perry, C. A., Annunziato, A. T. & Allis, C. D. *Proc. Natl Acad. Sci. USA* **92**, 1237–1241 (1995).
14. Candau, R. & Berger, S. L. *J. Biol. Chem.* **271**, 5237–5245 (1996).
15. Georgakopoulos, T., Gounalaki, N. & Thiros, G. *Mol. Gen. Genet.* **246**, 723–728 (1995).
16. Horiuchi, J., Silverman, N., Marcus, G. A. & Guarente, L. *Mol. Cell. Biol.* **15**, 1203–1209 (1995).
17. Yang, X.-J., Ogryzko, V. V., Nishikawa, J.-I., Howard, B. H. & Nakatani, Y. *Nature* **382**, 319–324 (1996).
18. Edmondson, D. G., Smith, M. M. & Roth, S. Y. *Genes Dev.* **10**, 1247–1259 (1996).
19. Gorovsky, M. A., Yao, M.-C., Keevert, J. B. & Pieger, G. L. *Methods Cell Biol.* **9**, 311–327 (1975).
20. Vavra, K. J., Allis, C. D. & Gorovsky, M. A. *J. Biol. Chem.* **257**, 2591–2598 (1982).
21. Braunstein, M., Sobel, R. E., Allis, C. D., Turner, B. M. & Broach, J. R. *Mol. Cell. Biol.* **16**, 4349–4356 (1996).
22. Horuchi, K. & Fujimoto, D. *Anal. Biochem.* **69**, 491–496 (1975).
23. Brownell, J. E. & Allis, C. D. *Proc. Natl Acad. Sci. USA* **92**, 6364–6368 (1995).

ACKNOWLEDGEMENTS. We thank O. Yarygina and C. Mizzen for the supply of yeast and chicken histones, respectively, J. Zhou for help in preparing bacterial extracts, and M. Hoyte for help in computer graphics. This work was supported by grants from the NIH.

CORRESPONDENCE and requests for materials should be addressed to C.D.A. (e-mail: cda@allis.biology.rochester.edu).

A new serine-protease fold revealed by the crystal structure of human cytomegalovirus protease

Liang Tong*, Chungeng Qian*, Marie-Josée Massariol†, Pierre R. Bonneau†, Michael G. Cordingley† & Lisette Lagacé†

*Boehringer Ingelheim Pharmaceuticals Inc., PO Box 368, 900 Ridgebury Road, Ridgefield, Connecticut 06877, USA

†Bio-Méga/Boehringer Ingelheim Research, Inc., 2100 rue Cunard, Laval, Québec H7S 2G5, Canada

HUMAN cytomegalovirus (hCMV), a herpesvirus, infects up to 70% of the general population in the United States and can cause morbidity and mortality in immunosuppressed individuals (organ-transplant recipients and AIDS patients) and congenitally infected newborns¹. hCMV protease is essential for the production of mature infectious virions, as it performs proteolytic processing near the carboxy terminus (M-site) of the viral assembly protein precursor (for a review, see ref. 2). hCMV protease is a serine protease², although it has little homology to other clans of serine proteases^{2,3}. Here we report the crystal structure of hCMV protease at 2.0 Å resolution, and show that it possesses a new polypeptide backbone fold. Ser 132 and His 63 are found in close proximity in the active site, confirming earlier biochemical and mutagenesis studies². The structure suggests that the third member of the triad is probably His 157. A dimer of the protease with an extensive interface is found in the crystal structure. This structure information will help in the design and optimization of inhibitors against herpesvirus proteases.

hCMV protease contains 256 amino-acid residues and shares significant sequence homology with the proteases of other herpesviruses². In addition to processing the assembly protein precursor, the protease also catalyses the cleavage at the R-site (residue 256) to release itself from its full-length gene product. Herpesvirus proteases prefer an Ala residue at P₁ and a Ser (or Ala) residue at P₁' (ref. 2). hCMV protease also has two internal cleavage sites (after residues 143 and 209). The hCMV protease

used here contains a mutation (A143Q) to disable one of these sites⁴. Limited processing at the other site was observed in solution, but the crystals contained only the intact protease.

The backbone fold of hCMV protease is different from those of the other clans³ of serine proteases with known structures (chymotrypsin, subtilisin and serine carboxypeptidase)^{5,6}, and represents a new fold for serine proteases. This also establishes the herpesvirus proteases as a new clan of serine proteases³. Moreover, an examination of the SCOP database⁷ failed to show any other structures with a similar fold, suggesting that the hCMV protease structure may represent the first observation of this protein fold. The structure consists of a seven-stranded, mostly antiparallel, β -barrel (β 1– β 7), which is surrounded by seven helices (α A– α G) (Fig. 1). Alternatively, the β -barrel can be described in terms of two β -sheets, with strands β 3, β 4, β 1 and β 7 in one sheet, and β 3, β 2, β 6, β 5 in the other. Strand β 3 switches between the two sheets on one side of the barrel. A short strand of three residues (β 8) is hydrogen-bonded to strand β 3, but it is not part of the β -barrel. The two β -sheets are arranged such that one opening of the β -barrel is rather narrow, whereas the other is more open. All the amino-acid side chains that point towards the inside of the β -barrel are hydrophobic and highly conserved² (see Supplementary Information). Of the helices, two (α B and α C) are located near the narrow end, and four (α D, α E, α F and α G) surround the middle of the barrel. Helix α A represents the few residues of the protein that are located near the open end of the barrel. Another hydrophobic core is formed by the packing of the β -barrel (strands β 3, β 4, β 1 and β 7) against helices α D, α E and α G. Residues involved in this core are also mostly conserved in sequence² (see Supplementary Information). The only crossover connection between the β -strands is right-handed, and is at the narrow end of the barrel. Five conserved regions that were identified on the basis of sequence comparisons of herpesvirus proteases² correspond mostly to the seven strands of the β -barrel. Therefore, the pattern of sequence conservation of herpesvirus proteases results from the structural requirement of maintaining the fold.

About 40 residues of hCMV protease are not observed in either molecule of the dimer in the current structure model. These are in four stretches: 1–8 at the amino terminus (loop L1), 46–52 (loop L3), 136–152 (loop L9), and 202–210 (loop L13). They correspond to regions of pronounced sequence variability among herpesvirus proteases² (see Supplementary Information). These residues are expected to be on the surface of the protein, and are likely to be flexible in structure. The two internal cleavage sites of hCMV protease are both located in the disordered stretches. It can be expected that their flexibility and surface location may make them accessible to attack by the protease.

As is common with the other protease families, we propose that all herpesvirus protease sequences be referred to by the hCMV protease sequence numbering. Site-directed mutagenesis and affinity-labelling studies have identified Ser 132 as the catalytic nucleophile of the herpesvirus protease². In the structure, Ser 132 is located in strand β 5 and is on the outside of the β -barrel. Residue His 63 (in loop L4) is located near Ser 132, in agreement with mutagenesis studies indicating that it is the second member of the triad². The other conserved histidine residue, His 157 (in strand β 6), which was identified to be important (although not essential) for catalytic activity by mutagenesis studies², is located near His 63. None of the Asp/Glu residues that have been proposed on the basis of mutagenesis studies² to be the third member of the triad are located near the His 63 residue. The structure therefore suggests that His 157 may be the third member of the catalytic triad in herpesvirus proteases. The relative positions of the Ser 132, His 63 and His 157 side chains in hCMV protease are similar to those found in other serine/cysteine proteases^{8,9}, especially the chymotrypsin family (Fig. 2). Superposition of Ser 132 and His 63 with the equivalent residues in chymotrypsin leads to a general overlap of the positions of His 157 in hCMV protease and Asp 102 in chymotrypsin, providing further support that His 157 may be the third member of the triad in

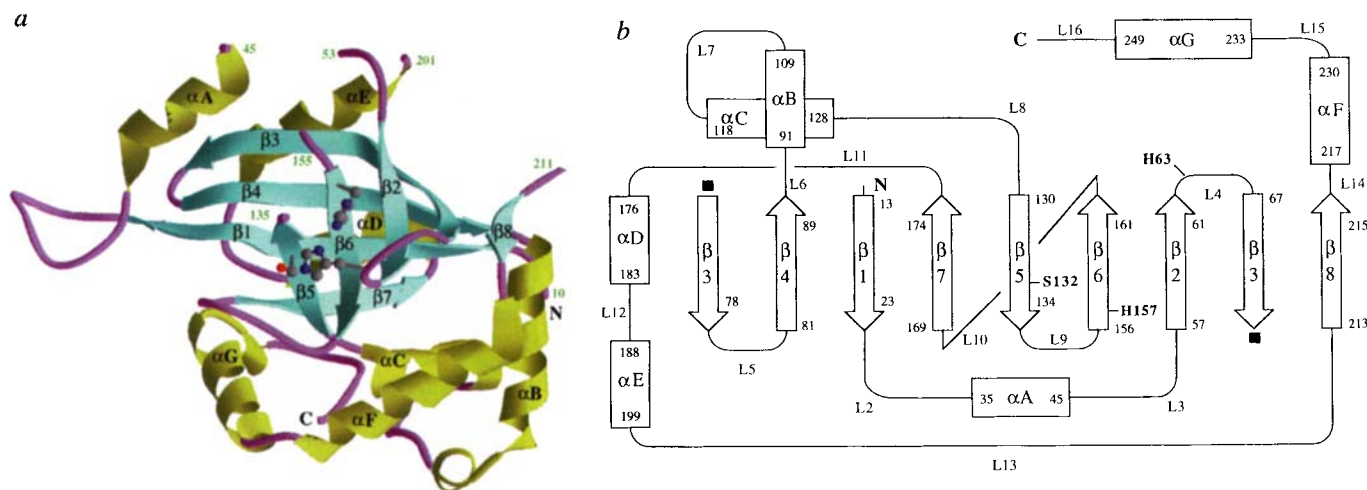


FIG. 1 *a*, Schematic drawing²² of the structure of hCMV protease. The β -strands are shown as arrowed ribbons in cyan, α -helices are yellow, and the connecting loops purple. The active-site residues (Ser 132, His 63 and His 157) are also shown. The secondary structure elements are labelled starting from the N terminus, the β -strands numerically ($\beta 1$ – $\beta 8$), and the α -helices alphabetically (αA – αG). The open end of the β -barrel is towards

the top left corner. *b*, Topological drawing of the backbone fold of hCMV protease. Strand $\beta 3$ is broken at the crossover position, and the break points are represented by black squares. The secondary structure elements are labelled. The loops are numbered from the N terminus (L1–L16). The location of Ser 132, His 63 and His 157 is also shown.

herpesvirus proteases. This is unique, as the other known triads generally contain an Asp, Glu or Asn residue as the third member. The catalytic activity of the H157D and H157E mutants, although about threefold greater than that of the H157A mutant, is still about five- to tenfold less than that of the wild-type enzyme (data not shown). Further studies will be needed to determine how His 157 functions as the third member of the catalytic triad.

The side-chain hydroxyl group of Ser 132 is partly disordered in this structure; its position (as observed here) may not correspond to the optimal position for catalysis. The distance between the hydroxyl group and the NE2 atom of His 63 is 3.3 Å, whereas that between the ND1 atom of His 63 and the NE2 atom of His 157 is 2.8 Å. The side chain of His 157 is mostly buried (accessible surface area, 10 Å²), whereas that of His 63 is partly exposed to solvent (accessible surface area, 70 Å²).

Despite the general similarity in the disposition of the catalytic residues, the difference in the backbone folds of the various enzymes means that the substrate-binding regions are located differently relative to the catalytic nucleophile. The oxyanion hole in hCMV protease is probably formed by the main-chain amide groups of residues 165 and 166. A water molecule is associated with the two groups in the current structure. The Ser 132 side chain is situated in a depression of the protein surface (Fig. 2). Residues in strands $\beta 5$ and $\beta 6$, helices αA , αF and αG , and loops L2, L4 and L15, are located near the active site and may contribute to the formation of the substrate-binding pockets. Residues in helix αB , from the other monomer of the protease dimer, are also located nearby. Residues Cys 161, Arg 165 and Arg 166 in this region are highly conserved among herpesvirus proteases. The side chain of Cys 161 is located next to that of Ser 132 and may be involved in the formation of the S_1 (or S'_1) binding pocket. Residues 136–152, disordered according to this crystal structure, could be near the active site and might contribute to the binding of substrates. The exact binding interactions of the substrate and the explanation of the substrate specificity of herpesvirus proteases will have to await the structure determination of substrate/inhibitor complexes.

Residues at the C terminus of the protease are located in a groove far from the active site and are covered with residues from loop L7, suggesting that there are probably conformational changes for the C-terminal residues after the cleavage at the R-site. The C-terminal carboxylate group forms a salt bridge with Lys 242, which is conserved to be a Lys or Arg in all herpesvirus

proteases² (see Supplementary Information). Lys 254 forms a salt bridge with Glu 122, which was proposed to be the third member of the triad². Tyr 253, strictly conserved among herpesvirus proteases, is situated in a hydrophobic pocket with the side-chain hydroxyl group forming a hydrogen bond with the carbonyl oxygen atom of residue 116 (loop L7). The structural importance of the Tyr 253 side chain may explain why this residue is conserved at the R-site of all herpesvirus proteases, but is never observed at the M-site².

Even though they are serine proteases, the activity of herpesvirus proteases is not easily inhibited by common serine protease inhibitors². However, Zn²⁺ ions, which normally inhibit cysteine proteases, can inhibit herpesvirus proteases at micromolar concentrations². X-ray diffraction data to 3.2 Å resolution were collected on a crystal of hCMV protease that was soaked overnight in 5 mM zinc sulphate. The difference electron-density map showed a binding site for Zn²⁺ in the protease, located near the side chains of His 63 and His 157. This suggests that zinc is an active-site inhibitor of herpesvirus proteases owing to the presence of two histidine residues in the catalytic triad.

The crystals of hCMV protease studied here contain dimers of the protease. Recent studies have shown that hCMV protease exists in a monomer–dimer equilibrium in solution, and that the active enzyme may be a dimer^{9,10}. The dimers observed here obey proper two-fold symmetry (with the exception of residues 25–46). The r.m.s. distance between 179 equivalent C α atoms of the two monomers is 0.28 Å. The dimer interface is formed mostly by helix αF packing against its symmetry-mate and helices αB and αC in the other monomer (Fig. 3). Helix αB has a kink near this interface, probably to prevent steric clashes with helix αF in the other monomer. The buried surface area per monomer is about 1,600 Å², suggesting that the interface is extensive. The residues at this interface show reduced variations among herpesvirus proteases² (see Supplementary Information). It may be expected that all herpesvirus proteases can form similar dimers.

The specific activity of herpesvirus proteases is rather low compared with that of other proteases². In the presence of 0.5 M Na₂SO₄, 50% glycerol or many other agents, the activity can be increased by 10- to 100-fold^{4,9–12}. Even with this increase, the activity is still much lower than that of the other proteases, perhaps because herpesvirus proteases have a weaker triad. To identify potential conformational changes in hCMV protease when Na₂SO₄ is introduced, crystals (originally in 0.4 M LiCl)

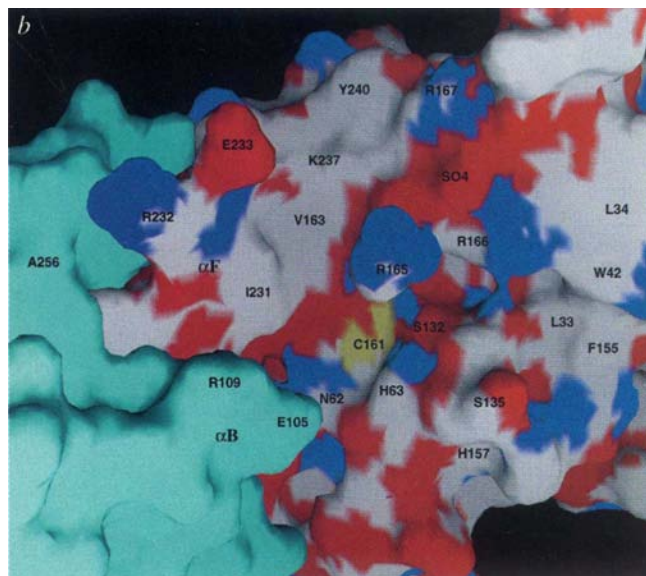
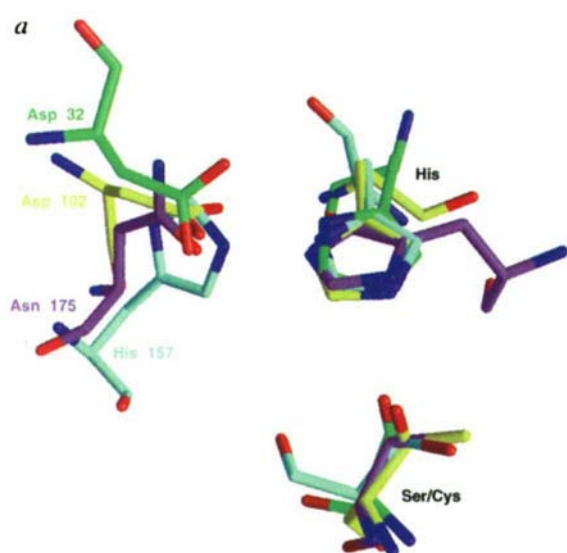


FIG. 2 *a*, Superposition of residues Ser 132, His 63 and His 157 in hCMV protease (cyan) with the catalytic triads of chymotrypsin (yellow), subtilisin (green) and papain (purple). The superposition was done manually, overlapping the C α atom of Ser 132 and the side-chain imidazole ring of His

63 with the equivalent atoms of the other triads. *b*, The molecular surface of hCMV protease near the active site with carbon atoms in white, oxygens in red, nitrogens in blue, and sulphur atoms in yellow for one monomer, and coloured cyan for the other monomer²³.

were exposed to 0.15 M Na₂SO₄ and a change in dimer organization was observed (Fig. 3). Changes in the conformations of individual residues may be possible as well, but the crystals in LiCl do not diffract well enough to allow a detailed structural comparison. The catalytic activity of hCMV protease increased about threefold by the change from 0.4 M LiCl to 0.15 M Na₂SO₄ (data not shown). However, with the current artificial mother liquor, the solubility of Na₂SO₄ is close to 0.15 M and a new set of conditions will be needed to look for any further changes in hCMV protease at higher concentrations of the salt.

The increase in the catalytic activity of herpesvirus proteases is due to a decrease in the K_m and/or an increase in the k_{cat} (refs 4, 9–12). It is possible that Na₂SO₄ and other agents stabilize the conformation of the loops that form the substrate-binding pockets and/or modify the shape of this binding region to facilitate catalysis. The observed change in the dimer organization in the presence of 0.15 M Na₂SO₄ may be related to this enhancement of catalytic activity. In addition, the disordered residues, especially 136–152, might become more ordered in the presence of higher concentrations of Na₂SO₄. These residues may also become more ordered on the binding of inhibitors in the active site. Structure determination of inhibitor complexes of hCMV protease is currently in progress. The structure we have determined, together with that from inhibitor complexes, will help to rationalize the structure–activity relationships of inhibitors, and will aid the design and optimization of new inhibitors against herpesvirus proteases. □

Methods

Crystal structure determination. The seleno-methionyl multiple-wavelength anomalous diffraction (MAD) technique¹³ was used, for which the introduction of two

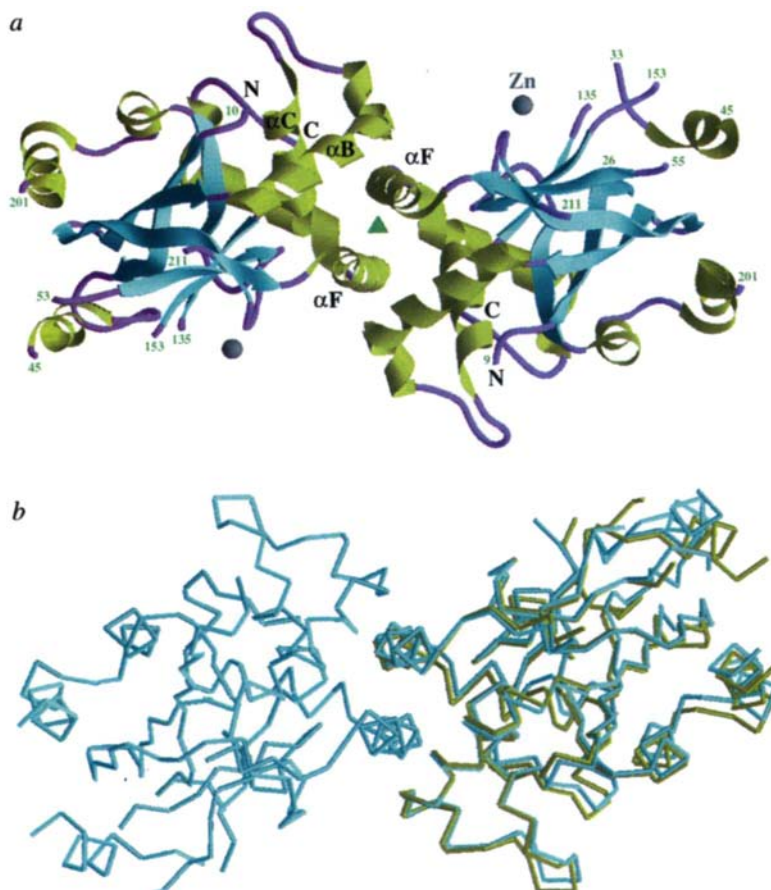


FIG. 3 *a*, The dimer of hCMV protease. The two-fold axis of the dimer is shown as a green triangle. The zinc ions are shown as grey spheres, marking the positions of the active sites. The dimer interface essentially involves residues in helix α F packing against helices α B, α C and α F in the other monomer. The view of this figure is roughly 90° away around the vertical axis from that of Fig. 1a. *b*, Superposition of the C α trace of the dimer in 0.15 M Na₂SO₄ (cyan) with that in 0.4 M LiCl (yellow). The superposition was done with residues from one monomer only.

additional Met residues by mutagenesis (T181M, L229M) was crucial. Non-crystallographic symmetry averaging (among crystal forms) was essential to improve the original MAD phases.

Crystallization. hCMV protease was expressed in *Escherichia coli* and purified by anion- and cation-exchange chromatography. Initial crystallization condition was found by the sparse-matrix sampling method¹⁴, with a commercial kit (Hampton). Purified protease was concentrated to about 25 mg ml⁻¹ in 20 mM sodium acetate (pH 5.0), 80 mM NaCl and 1 mM DTT. The reservoir solution contained 16% PEG 4000, 0.1 M MES (pH 6.0), 0.4 M LiCl, 10% glycerol and 5% *t*-butanol. The crystals were grown at room temperature and transferred in a few steps to an artificial mother liquor containing 30% PEG.

Data collection. Diffraction data were collected at cryo-temperature on an R-axis imaging plate system mounted on a Rigaku rotating anode generator, on Fuji imaging plates at the X4A beamline, and on a MAR detector at the X25 beamline at the Brookhaven National Laboratory. Diffraction images were processed using DENZO¹⁵. The crystals belonged to space group *P*4₂2₁ but exhibited significant differences in cell parameters (called 'crystal forms' here). There is a dimer of the protease in the asymmetric unit, with significant variations in its orientation and position in the different crystal forms.

MAD phasing. Initial phase information was obtained with the MADSYS package¹³ based on the MAD data (at 3 Å resolution) for the L229M mutant collected at the X4A beamline (crystal form A, *a* = *b* = 70.1 Å, *c* = 215.4 Å). The Se positions were determined from *F*_A Patterson maps using the program PATSOL¹⁶. The phase information was used to solve the native crystal forms B (*a* = 68.5 Å, *c* = 211.5 Å) and C (72.7 Å, 215.3 Å) by molecular replacement with the REPLACE suite¹⁷. Subsequent averaging among the three crystal forms, with a locally written program (L.T., unpublished), produced an electron-density map with recognizable secondary structure elements. The Se and Hg sites, identified from a difference electron-density map for a K₂Hg(SCN)₄ soak, helped in the development of a preliminary trace and an initial model was built with the program FRODO¹⁸. It was discovered that replacement of the LiCl component in the artificial mother liquor with 0.15 M Na₂SO₄ produced a dramatic improvement (from 3 Å to 2 Å resolution) in the diffraction quality of the crystals. A MAD data set to 2.0 Å resolution was collected on such a crystal of the T181M/L229M mutant at the X25 beamline (crystal form D, *a* = 71.0 Å, *c* = 209.3 Å). The phasing statistics at 2.5 Å resolution based on the data, with either the MADSYS package or X-PLOR¹⁹, were rather poor. However, after two-fold averaging at 2.5 Å resolution, an excellent electron-density map was obtained showing amino-acid side chains and the main-chain carbonyl groups. The trace developed earlier was found to be essentially correct. Only about 190 residues could be located in the averaged electron-density map. Residues 35–45 (helix α A) were found after partial model phase combination with the SIGMAA program²⁰.

Structure refinement. The structure refinement was performed using the program X-PLOR¹⁹, for reflections between 6.0 and 2.0 Å resolution with *R* > 2 σ . NCS restraints were used at the early stages of refinement. Residues 26–34 in one monomer and some other missing residues were located in the 2*F*_o – *F*_c electron-density maps. The current structure model contains 419 residues of the dimer, one sulphate ion, and 249 water molecules. The *R* factor is 22.1% for 30,822 reflections between 6.0 and 2.0 Å resolution (90% complete). The free *R* factor²¹, for 7.5% of the reflections, is 28.8%. The r.m.s. deviation in bond lengths is 0.010 Å, and that in bond angles is 1.8°. Details of the cloning, expression, purification, crystallization and structure determination of hCMV protease will be presented elsewhere.

SUPPLEMENTARY INFORMATION. Available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from Mary Sheehan at the London editorial office of Nature.

ACKNOWLEDGEMENTS. We thank H. Wu for discussions and help with data collection at the Brookhaven National Laboratory; C. Ogata and D. Cook at the X4A beamline, and L. Berman, M. Capel and Z. Yin at the X25 beamline, for help with the data collection; W. Hendrickson for the MADSYS package; A. Br nger for a pre-release version of X-PLOR 4.0; R. Read for the program SIGMAA; T. Peat for competent *E. coli* DL41 cells; W. Davidson for characterization of the protein samples by mass spectrometry; and S. Ghaemmaghami for help with the production of seleno-methionyl hCMV protease mutants.

CORRESPONDENCE and requests for materials should be addressed to L.T. (e-mail: b4w@cc.purdue.edu). The atomic coordinates have been deposited in the Brookhaven Protein Data Bank (accession number 1WP0).

Unique fold and active site in cytomegalovirus protease

Xiayang Qiu, Jeffrey S. Culp*, Anthony G. DiLella†, Brian Hellmig*, Susan S. Hoog, Cheryl A. Janson*, Ward W. Smith & Sherin S. Abdel-Meguid

Departments of Macromolecular Sciences, * Protein Biochemistry, and † Molecular Genetics, SmithKline Beecham Pharmaceuticals, 709 Swedeland Road, King of Prussia, Pennsylvania 19406, USA

HUMAN herpesviruses are responsible for a variety of diseases. They are divided into three subfamilies: alpha includes herpes simplex viruses (HSV-1 and HSV-2) and varicella-zoster virus (VZV); beta includes cytomegalovirus (CMV) and human herpesvirus-6 (HHV-6); and gamma includes Epstein–Barr virus (EBV). Each virus encodes a serine protease that is essential for its replication^{1–14} and is a potential target for therapeutic intervention. Human CMV is a ubiquitous opportunistic pathogen that can result in life-threatening infections in congenitally infected infants, immunocompromised individuals and immunosuppressed cancer or transplant patients¹⁵. Here we report the crystal structure of human CMV protease at 2.5 Å resolution. The structure reveals a fold that has not been reported for any other serine protease, and an active site consisting of a novel catalytic triad in which the third member is a histidine instead of an aspartic acid, or possibly a catalytic tetrad consisting of a serine, two histidines and an aspartic acid. An unusual dimer interface that is important to the protease activity has also been identified.

The herpesvirus proteases are encoded as precursor proteins that catalyse their own cleavage to produce a fully active amino-terminal domain of relative molecular mass ~28,000 (*M*_r 28K)^{3–6}. These protease domains show considerable sequence homology: about 30% identity between members of different subfamilies, and as much as 90% identity within the same subfamily⁶. There is little sequence homology to other known proteins, including the absence of the conserved G-X-S/C-G-G sequence for chymotrypsin-like, and G-T-S-M/A for subtilisin-like, proteases. The herpes proteases have similar substrate specificity: they all cleave a peptide bond between an alanine and a serine⁵. Most common covalent serine protease inhibitors, such as phenylmethylsulphonyl fluoride (PMSF) and N-tosyl-L-leucine chloromethyl ketone (TLCK), are not good inhibitors of these proteases^{5–10}. It is therefore reasonable to assume that all herpes proteases share a common fold, and it is not surprising that the CMV protease crystal structure reveals a fold and an active site that are distinct from those of other serine proteases.

Unlike classical serine proteases, which have two distinct β -barrel domains, the structure of CMV protease has a single domain. The overall fold can be best described as a seven-stranded β -barrel core with seven α -helices on three sides (Fig. 1*a*). The core β -barrel is orthogonally packed¹⁰ but has two distinct features.

Received 9 May; accepted 25 July 1996.

- Fields, B. N. *et al.* (eds) *Virology* Vol. 2, ch. 64–73 (Raven, New York, 1990).
- Gibson, W., Welch, A. R. & Hall, M. R. T. *Perspect. Drug Discov. Design* **2**, 413–426 (1995).
- Rawlings, N. D. & Barrett, A. J. *Methods Enzymol.* **244**, 19–61 (1994).
- Pinko, C. *et al.* *J. Biol. Chem.* **270**, 23634–23640 (1995).
- Ollis, D. L. *et al.* *Protein Eng.* **5**, 197–211 (1992).
- Perona, J. J. & Craik, C. S. *Protein Sci.* **4**, 337–360 (1995).
- Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. *J. Mol. Biol.* **247**, 536–540 (1995).
- Yamamoto, A. *et al.* *Biochemistry* **31**, 11305–11309 (1992).
- Darke, P. L. *et al.* *J. Biol. Chem.* **271**, 7445–7449 (1996).
- Margosiak, S. A., Vanderpool, D. L., Sisson, W., Pinko, C. & Kan, C.-C. *Biochemistry* **35**, 5300–5307 (1996).
- Burck, P. J. *et al.* *J. Virol.* **68**, 2937–2946 (1994).
- Hall, D. L. & Darke, P. L. *J. Biol. Chem.* **270**, 22697–22700 (1995).
- Hendrickson, W. A. *Science* **254**, 51–58 (1991).
- Jancarik, J. & Kim, S.-H. *J. Appl. Crystallogr.* **24**, 409–411 (1991).
- Otwinski, Z. In *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Laboratory, Daresbury, 1993).
- Tong, L. & Rossmann, M. G. *J. Appl. Crystallogr.* **26**, 15–21 (1993).
- Tong, L. *J. Appl. Crystallogr.* **26**, 748–751 (1993).
- Jones, T. A. *J. Appl. Crystallogr.* **11**, 268–272 (1978).
- Br nger, A. T. *The X-PLOR Manual*, Version 3.0 (Yale University, New Haven, CT, 1992).
- Read, R. J. *Acta Crystallogr.* **A42**, 140–149 (1986).
- Br nger, A. T. *Nature* **355**, 472–475 (1992).
- Carson, M. J. *Mol. Graph.* **5**, 103–106 (1987).
- Nicholls, A., Sharp, K. A. & Honig, B. *Proteins* **11**, 281–296 (1991).

First, it contains two parallel strands (Fig. 1b), whereas a standard orthogonally packed β -barrel is exclusively antiparallel¹⁶. Second, strand B3 is a β -bend that closes one corner of the barrel, but the other corner lacks this kind of classical closure and is maintained by two hydrogen bonds between strands B5 and B7. The N-terminal domain of trypsin is also an orthogonally packed β -barrel, but the superposition of the two barrels did not reveal

much resemblance, and showed that the active sites are positioned in different regions of the fold. The first four strands of CMV protease (B1–B4) form a Greek key motif (Fig. 1b) quite different from that of trypsin. It is reasonable to conclude that the CMV protease β -barrel is evolutionarily unrelated to other serine protease β -barrels, and that the CMV protease fold is unique among known serine proteases.

An unusual dimer interface has been identified between two CMV protease monomers related by two-fold crystallographic axes (Fig. 2). The dimer interface contains four helices (A1, A2, A3 and A6) of one monomer surrounding helix A6 of the other, with the two symmetry-related A6 helices being almost parallel (Fig. 2). The dimer interface is predominantly hydrophobic and conserved among the herpes proteases, with an interface area of about 1,300 Å². CMV protease is thought to be active as a dimer^{17,18}, although from the crystal structure it is unclear why the dimer is required for activity, as the dimer interface is not in the immediate vicinity of the active site, and the active sites of the monomers are a long way apart (Fig. 2). However, rearrangements of the helices at the interface in the absence of a dimer could have profound effects on the conformation in the active-site region. For example, helix A6 might move towards the active site (Fig. 2), affecting the positioning of active-site residues or blocking access to the substrate.

The catalytic mechanism of classical serine proteases involves an active-site triad composed of a serine, a histidine and an aspartic acid. The crystal structure of human CMV protease reveals an active site containing a serine (Ser 132) and a histidine (His 63), with the third catalytic residue being a histidine (His 157) instead of an aspartic acid (Fig. 3). The OG atom of Ser 132 is in the vicinity of His 63, and is 3.3 Å from its NE₂ nitrogen. The conserved second histidine (His 157) is connected by a hydrogen bond to His 63 (His 63 ND₁–His 157 NE₂, 3.2 Å), making it the third member of the catalytic triad. The only acidic residue in the vicinity is Asp 65, with its OD₁ atom 3.9 Å away from the ND₁ nitrogen of His 157. In the crystal structure, Asp 65 forms a salt bridge with Arg 109 of a neighbouring molecule other than the dimer. In the absence of this salt bridge in solution, the Asp-65 side chain could move to form a hydrogen bond with His 157 and act as a proton acceptor in a 'catalytic tetrad'. Mutagenesis and chemical modification studies had identified Ser 132 and His 63 as being directly involved in catalysis, but failed to identify the third member of the triad^{5–8}. None of the acidic residues is conserved among all herpes proteases. Glu 122 has been proposed to be a member of the catalytic triad⁷, but is 25 Å away from the catalytic site in the structure. From our structure, we conclude that the active site of CMV protease consists of either a novel triad of Ser

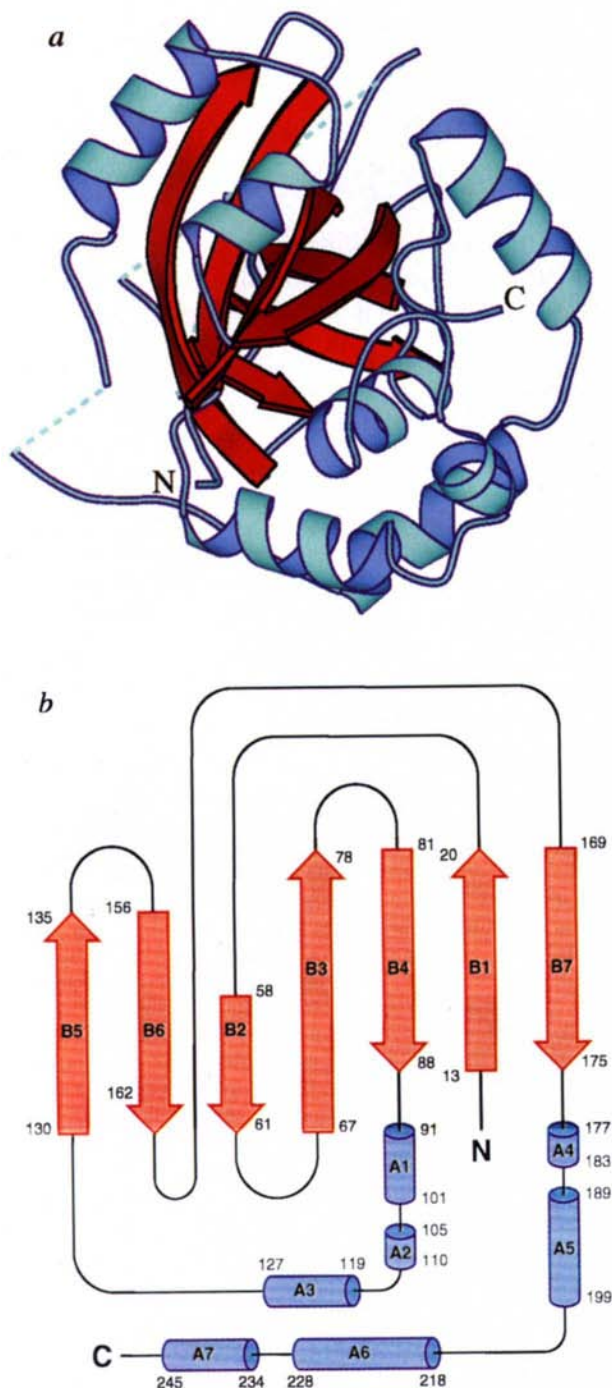


FIG. 1 The CMV protease monomer. *a*, Ribbon diagram of the structure with the core β -barrel highlighted in red. The N and C termini are indicated. Disordered portions of the structure are represented by broken lines. The diagram was drawn with the program MOLSCRIPT²⁷. *b*, Topology diagram of the monomer with helices represented as cylinders and strands as arrows. Strands B5 and B7 are adjacent in the structure.



FIG. 2 The CMV protease dimer viewed parallel to the crystallographic two-fold axis. Subunits are drawn in red and blue, with residues of the catalytic triad in yellow. The two parallel helices are indicated by A6. The diagram was drawn with the program MOLSCRIPT²⁷.

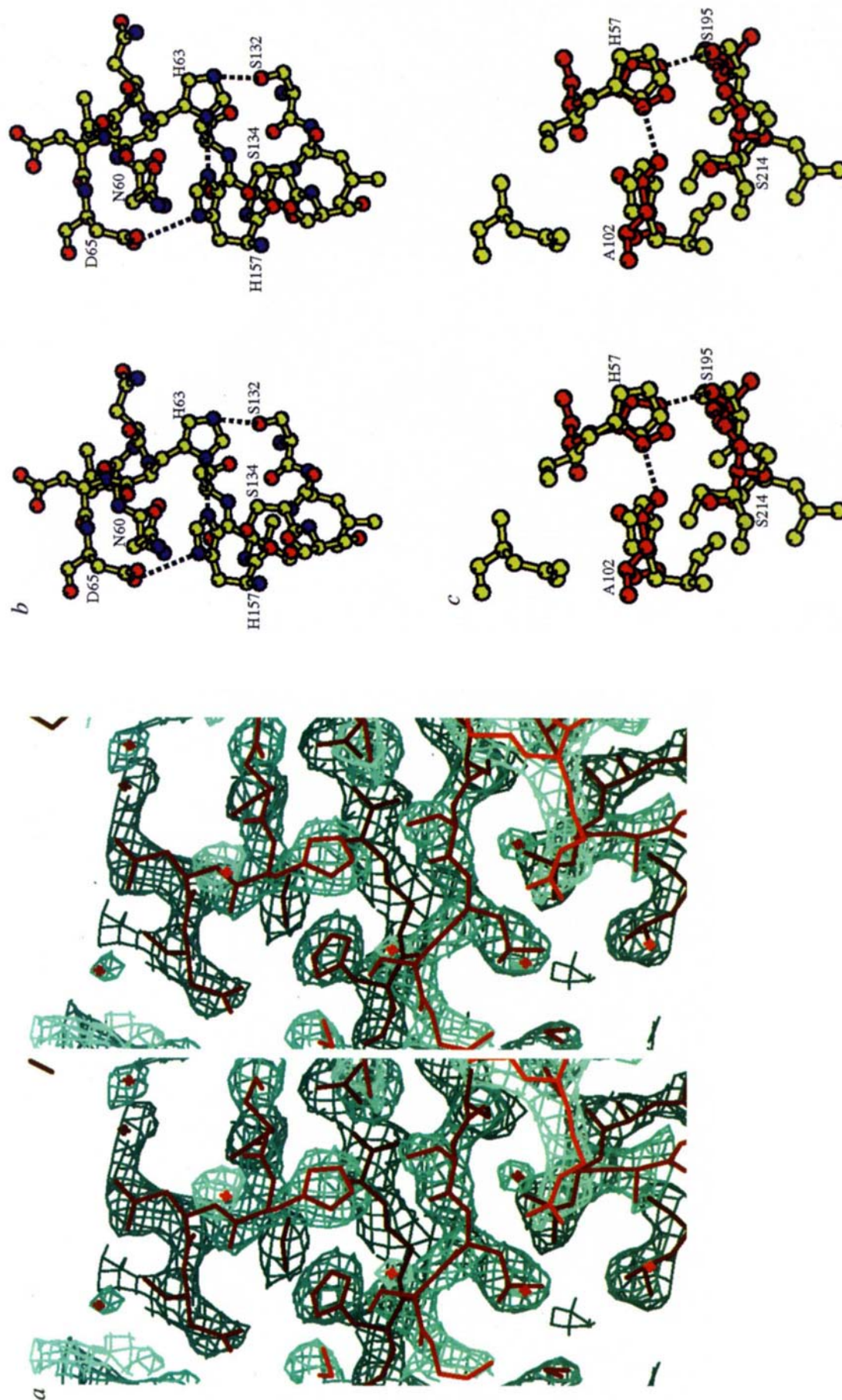


FIG. 3 The active site of CMV protease. *a*, Stereo view of the final $2F_o - F_c$ electron-density map contoured at 1.2σ . *b*, Stereoview of the CMV protease active site in ball-and-stick representation drawn with MOLSCRIPT²⁷. Carbon atoms are yellow, nitrogen blue, and oxygen red. The catalytic residues are S 132, H 63, H 157 and possibly D 65. The postulated 'proton transfer pathway' is shown by broken lines. *c*, Stereoview of the superposition between the active site of CMV protease and that of the classical serine protease trypsin. The r.m.s. difference between comparable atoms of the two triads is 1.6 Å. The CMV protease is shown in yellow. Trypsin is shown in red and labelled. The 'proton transfer pathway' in trypsin is depicted by broken lines.

TABLE 1 Statistics of structure determination

Data set	Resolution (Å)	Observed (>1σ)	Unique (>1σ)	Complete (%)	R _m (%)	R _{iso} (%)	Phasing power	R _{Cullis}
Native	2.5	22,880	7,644	90	6.0	—	—	—
MeHgCl	3.2	17,566	3,932	97	10.1	28.0	2.8	0.50
MeHgCl	3.5	3,698	1,958	76	19.8	36.5	1.4	0.75
UO ₂ Ac ₂	3.8	9,122	2,330	99	23.2	27.5	1.1	0.83
BkrHg ₂	3.2	14,867	4,425	97	9.8	30.7	1.9	0.68
LuCl ₃	3.2	9,984	3,734	95	10.8	23.4	2.0	0.63
K ₂ PtCl ₄	4.1	3,017	1,673	79	13.1	25.9	1.4	0.77
SmCl ₃	3.7	12,708	2,804	98	13.7	26.3	2.4	0.54

MIRAS figure of merit (30–3.2 Å): 0.70

XPLOR refinement								
Resolution		7.0–2.5 Å		reflections (>1σ)			7,193	
Protein atoms		1,504		solvent			73	
R.m.s. bond length		0.016 Å		r.m.s. bond angle			2.2°	
R-factor		0.185						

$R_m = \sum |I_i - \langle I \rangle| / \sum I_i$; $R_{iso} = \sum |F_{PH} - F_P| / \sum F_P$; phasing power = r.m.s. isomorphous difference/r.m.s. residual lack of closure; $R_{Cullis} = \sum |F_{H_{obs}} - F_{H_{cal}}| / \sum |F_{H_{obs}}|$ defined for centric reflections; R-factor = $\sum ||F_o| - |F_c|| / \sum |F_o|$.

132, His 63 and His 157, or a unique tetrad consisting of Ser 132, His 63, His 157 and Asp 65 (Fig. 3b). In the 'catalytic tetrad', His 157 would act as an extra component in a novel 'relay' proton-transfer mechanism. The lack of sequence conservation for Asp 65 raises a serious concern regarding the validity of the tetrad hypothesis. It is possible that Asp 65 is substituted by an acidic residue from a different segment of sequence in other herpes proteases; this will become clear when structures of other herpesviruses proteases are known.

Overlay of Ser 132 and His 63 of CMV protease on Ser 195 and His 57 of trypsin puts His 157 on top of Asp 102 of trypsin (Fig. 3c). This supports the role of His 157 in catalysis, which is to stabilize the conformation of His 63 and possibly delocalize a proton¹⁹. Despite different tertiary structures of the two enzymes, many other conservations can be identified in the same overlay. Another example of this apparent convergent evolution is that Ser 134 appears to occupy an identical position to Ser 214 of trypsin. Ser 134 of CMV protease interacts strongly with His 157 (Ser 134 OG-His 157 NE₂, 2.6 Å). In trypsin, Ser 214 also forms a hydrogen bond with Asp 102. However, Ser 134 is not essential to the protease activity, as shown by mutagenesis studies⁶. The overlay also allowed us to identify an oxyanion hole for CMV protease. In trypsin, an oxyanion is held by backbone nitrogens of Gly 193 and Ser 195. In CMV protease, the backbone nitrogen of Arg 165 is at a position identical to trypsin Gly 193 N. Also found in the vicinity is a water molecule held by the side chain of Arg 166 which interacts with Leu 20 and Leu 133 (H₂O–Arg 166 NH₁, 2.7 Å; H₂O–Arg 166 NZ, 3.2 Å; H₂O–Leu 20 O, 2.6 Å; H₂O–Leu 133 N, 3.0 Å). We are unable to determine whether the oxyanion is held only by Arg 165 N or by Arg 165 N and H₂O. Because these residues (20, 133, 165 and 166) are conserved among herpes proteases⁶, we think that the oxyanion pocket of CMV protease is in this general region.

A better understanding of the structural basis for the substrate specificity of the herpes proteases must await the structure determination of enzyme–ligand complexes, but the current structure provides some initial clues. The active site of human CMV protease sits at a very shallow and mostly exposed region of the protease (Fig. 2). Such shallowness is reasonable, given that the scissile bond recognized by all herpes proteases is between two small residues (Ala–Ser). The cavity comprising the S' subsites is delineated by part of helix A6, strand B6, loop 162–169 and loop 228–234 (Fig. 1b). This region of the active-site cavity is highly reminiscent of that of classical serine proteases¹⁹. The cavity comprising the S subsites, however, is much more difficult to delineate. It is formed partly by strand B5 and probably residues

from two missing surface loops (25–55 and 143–153). The importance of these two loops for substrate binding can also be inferred from mutagenesis studies. A five-residue deletion around 143 and a Glu 31 Ala mutant were shown to alter CMV protease substrate specificity⁶. Given the proximity of these loops to the active-site cavity, we speculate that they are important to substrate recognition, and probably become ordered on ligand binding. □

Methods

The enzyme used in these studies was mutated at position 143 from an alanine to a valine. This mutation had no effect on the enzyme activity, but was necessary to eliminate the cleavage after Ala 143 seen in preparations of the native enzyme⁶. Crystals were grown using the sitting-drop method of vapour diffusion from a well containing 30% PEG400 in 0.1 M phosphate–citrate buffer, pH 4, and a drop of tetragonal 7 mg ml^{−1} protease mixed 1:1 with the well solution. The crystals belong to the space group P4₃22, with *a* = *b* = 58.7 Å, *c* = 131.0 Å. There is one molecule per asymmetric unit and ~40% solvent content. The native data were measured at the CHESS A-1 beamline using a CCD (charge-coupled device) detector and processed with the programs DENZO/SCALEPACK²⁰. Derivative data sets were collected with a Siemens area detector on a Rigaku CuK_α source and processed with XENGEN²¹. The structure was determined using multiple isomorphous replacement methods (Table 1). Heavy-atom derivatives were prepared by soaking the crystals for 1–4 days in saturated methyl-HgCl (pH 4 or 5), saturated Baker's dimercury, 1 mM UO₂Ac₂, 1 mM K₂PtCl₄, 0.5 mM LuCl₃ or 0.5 mM SmCl₃. Heavy-atom positions were identified by difference Patterson and difference Fourier methods using the programs in CCP4 (ref. 22). Anomalous signals from the derivatives allowed the determination of the space group's chirality and heavy-atom coordinates. Heavy-atom refinement and phasing were performed using MLPHARE²³, and the resulting MIRAS map was interpretable. After density modification using solvent flattening, histogram matching and skeletonization, 75% of the molecules (most with side chains) were built in the computer graphics program XTALVIEW²⁴. The R-factor of the starting model was 43.8% (10–3.0 Å); it decreased quickly to 28.3% after only 200 cycles of XPLOR²⁵ positional refinement. His 63 and His 157 were refined as carrying a single proton at their ND1 atoms. Although human CMV protease contains 256 amino acids, our current model is represented by 202 residues. Residues 1–8 and residues in three surface loops (25–55, 143–153 and 205–208), which are not conserved among herpes proteases, were disordered in the crystal. The remaining residues were fitted in well-defined electron density (Fig. 3a). The final model was refined at 2.5 Å to an R-factor of 18.5% (Table 1). The program PROCHECK²⁶ was used to check the stereochemical and geometrical outliers in the final structure, and all the results are of good quality and within allowed regions.

Received 20 May; accepted 25 July 1996.

1. Liu, F. & Roizman, B. *J. Virol.* **65**, 5149–5156 (1991).
2. Gao, M. et al. *J. Virol.* **68**, 3702–3712 (1994).
3. Liu, F. & Roizman, B. *J. Virol.* **67**, 1300–1309 (1993).
4. Weinheimer et al. *J. Virol.* **67**, 5813–5822 (1993).
5. Liu, F. & Roizman, B. *Proc. Natl Acad. Sci. USA* **89**, 2076–2080 (1992).

6. Welch, A. R., McNally, L. M., Hall, M. R. T. & Gibson, W. J. *Virology* **67**, 7360–7372 (1993).
7. Cox, G. A., Wakulchik, M., Sassmannshausen, L. M., Gibson, W. & Villarreal, E. C. *J. Virology* **69**, 4524–4528 (1995).
8. Stevens, J. T. et al. *J. Biochem.* **226**, 361–367 (1994).
9. Burck, P. J. et al. *J. Virology* **68**, 2937–2946 (1994).
10. Holwerda, B. C. et al. *J. Biol. Chem.* **269**, 25911–25915 (1994).
11. Baum, E. Z. et al. *J. Virology* **67**, 497–506 (1993).
12. O'Boyle II, D. R., Wager-Smith, K., Stevens III, J. T. & Weinheimer, S. P. *J. Biol. Chem.* **270**, 4753–4758 (1993).
13. Welch, A. R., Villarreal, E. C. & Gibson, W. J. *Virology* **69**, 341–347 (1993).
14. Dilanni, C. L. et al. *J. Biol. Chem.* **268**, 25449–25454 (1993).
15. Alford, C. A. & Britt, W. J. in *Virology* 2nd edn (eds Knipe, D. M. and Fields, B. N.) 1981–2010 (Raven, New York, 1990).
16. Chothia, C. & Janin, J. *Biochemistry* **21**, 3955–3965 (1982).
17. Margosiak, S. A., Vanderpool, D. L., Sisson, W., Pinko, C. & Kan, C. *Biochemistry* **35**, 5300–5307 (1995).
18. Darke, P. L. et al. *J. Biol. Chem.* **271**, 7445–7449 (1996).
19. Perona, J. J. & Craik, C. S. *Protein Sci.* **4**, 337–360 (1995).
20. Otwinowski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (Daresbury Laboratory, Warrington, 1993).
21. Howard, A. J. et al. *J. Appl. Crystallogr.* **A47**, 110–119 (1994).
22. Collaborative Computational Project *Acta Crystallogr.* **D50**, 760–763 (1994).
23. Otwinowski, Z. in *Isomorphous Replacement and Anomalous Scattering* (eds Wolff, W., Evans, P. R. & Leslie, A. G. W.) 80–86 (Daresbury Laboratory, Warrington, 1991).
24. McKee, D. E. *Practical Protein Crystallography* (Academic, San Diego, 1993).
25. Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
26. Laskowski, R. A., MacArthur, M. W., Morris, A. L. & Thornton, J. M. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
27. Kraulis, P. J. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

ACKNOWLEDGEMENTS. We thank L. Gorniak and A. Patel for activity assays; K. O'Donnell and D. O'Shanney for contributions to early stages of purification; C. Debouck, G. Glover, R. Jarvest and M. Rosenberg for encouragement and support; and R. Jarvest for comments on the manuscript.

CORRESPONDENCE and requests for materials should be addressed to S. S. Abdel-Meguid (e-mail: sherin_s.abdel-meguid@sbbnrd.com). The coordinates have been deposited in the Brookhaven Protein Databank, accession no. 1LAY, where they will be held for one year before release.

Three-dimensional structure of human cytomegalovirus protease

Huey-Sheng Shieh*, Ravi G. Kurumbail*, Anna M. Stevens*, Roderick A. Stegeman*, Eric J. Sturman*, Jina Y. Pak*, Arthur J. Wittwer†, Mark O. Palmier‡, Roger C. Wiegand†, Barry C. Holwerda† & William C. Stallings*

Departments of * Medicinal and Structural Chemistry, † Infectious Diseases, and ‡ Biological Sciences, Monsanto/Searle, Searle Discovery Research, 700 Chesterfield Parkway North, St Louis, Missouri 63198, USA

HERPESVIRUSES encode a serine protease^{1,2} that specifically cleaves assembly protein³. This protease is critical for replication⁴, and represents a new target for antiviral drug design⁵. Here we report the three-dimensional structure of the protease from human cytomegalovirus (hCMV) at 2.27 Å resolution. The structure reveals a unique fold and new catalytic strategy for cleavage. The monomer fold of the enzyme, a seven-stranded β -barrel encircled by a chain of helices that form the carboxy terminus of the molecule, is unrelated to those observed in classic serine proteases such as chymotrypsin and subtilisin. The serine nucleophile at position 132 is activated by two juxtaposed histidine residues at positions 63 and 157. Dimerization, which seems to be necessary for activity^{6,7}, is observed in the crystals. Correlations of the structure with the sequences of herpesvirus proteases^{1,5,8} suggest that dimerization may confer specificity and recognition in substrate binding.

Infection by human herpesviruses is an important cause of new and recurrent diseases in all populations. The nuclei of cells infected by herpesviruses typically contain immature capsid forms, B-capsids, that have a phosphoprotein called the assembly protein³, of relative molecular mass 35,000 to 40,000 (M_r 35K–40K). Herpesviruses also encode a serine protease, which cleaves the assembly protein precursor at a site near its C terminus², a

process necessary for viral replication⁴. The protease and assembly protein are encoded by the same open reading frame in the viral genome. Consequently, the protease is synthesized in infected cells as the proximal component of a fusion protein with the assembly protein. Most of the assembly protein, however, is synthesized alone from a separate internal promoter. The protease recognizes and cleaves at a consensus sequence, (V,L)-(K,E,N)-A/(S,A), found at the 'release site' between protease and assembly protein in the fusion, and at the C-terminal 'maturation site' of the assembly protein. Processing at the assembly protein maturation site may be the main function of the protease in viral capsid maturation.

We have determined the crystal structure of hCMV protease at 2.27 Å resolution to study the active site of herpesvirus proteases and the effect of dimerization of the subunits. Unusually⁹, the monomer fold has a seven-stranded β -barrel (Fig. 1a) and most of the strands are antiparallel. The β -barrel begins at residue 14, near the first residue visible in electron-density maps, and ends at residue 174. The connection between strands 2 and 6 (residues 89–127), located on one of the open ends of the barrel, is well ordered with three helical regions (helices 3–5). However, electron density for some of the more extended interstrand connections located on the opposite end of the barrel cannot be interpreted with confidence. Beyond residue 174, the structure of the C-terminal residues has four helical segments (helices 6–9), but electron density for residues 202–211 is discontinuous. Although residues 175–201 fold as two antiparallel helices, the four helices do not otherwise assemble into a compact helical domain. Instead, the C terminus is just long enough to completely encircle the N-terminal β -barrel (Fig. 1b), resulting in a novel protein fold which makes it the first of a new class of proteases. The surrounding helices encase the β -barrel, but leave exposed the face containing the active site.

The active site of herpesvirus proteases features a conserved serine residue, at position 132 in hCMV protease and 129 in the herpes simplex virus type-1 enzyme, that has been identified as the catalytic nucleophile by mutagenesis and covalent modification with diisopropyl fluorophosphate^{10–12}. Mutagenesis studies¹¹ also showed that there is a conserved histidine, His 63 in hCMV, which is necessary for the activity. N^ε of the imidazole side chain of His 63 forms a hydrogen bond to the Ser 132 hydroxyl group. Other serine proteases have a conserved aspartyl residue that forms a third member of a well recognized catalytic triad. No aspartic or glutamic acid residues form side-chain neighbours with His 63, nor are they conserved in herpesvirus protease sequences. Instead, N^δ of His 63 forms a hydrogen bond to N^ε of His 157, which is the only other conserved histidine residue in herpesvirus proteases (Fig. 2). The structure is consistent with data¹¹ suggesting that His 63 functions as a general base that activates Ser 132 during catalysis. The second histidine residue, His 157, may increase the pK of His 63 by forming an additional hydrogen bond to its imidazole ring. So His 157 is the third member of the catalytic triad, and is a surrogate for the acidic member of the classic serine protease triad. His 157 is not absolutely required for catalysis, as a His 157 variant is competent to cleave assembly protein¹¹, but kinetic parameters for this variant enzyme have not been reported.

All serine proteases have a site that stabilizes the binding of anionic tetrahedral intermediates during catalysis. In the chymotrypsin class, this is the oxyanion hole, formed by backbone N–H groups of residues 193 and 195. In the subtilisin family, the oxyanion hole is formed by the side chain of Asn 155 and the backbone N–H of the serine nucleophile. The conserved arginine residues at positions 165 and 166 are close to Ser 132 (Fig. 2), and so these residues might form the oxyanion hole in herpesvirus proteases. We have superimposed the C α atoms of Ser 132, His 63 and His 157 on the C α atoms of Ser 195, His 57 and Asp 102, respectively, of trypsin in the structure of its complex with bovine pancreatic trypsin inhibitor¹³; similarly, hCMV protease was also superimposed on the complex of subtilisin with the inhibitor, eglin¹⁴. The superposition (not shown) correlates the backbone

FIG. 1 *a*, Schematic chain trace of the hCMV protease monomer. Strands are labelled S1–S7, helices H1–H9. Connections not visible in electron-density maps are indicated by a broken line. The topology of the β -barrel is shown, but the encirclement of the C-terminal helices about the barrel is not represented. Neighbouring strands 4 and 5 of the β -barrel are parallel, but the others are antiparallel; strands 6 and 7 are connected by only two interstrand hydrogen bonds. Not all of the residues of the barrel are in purely β -conformations; for example, residues 20 and 73 are involved in well-defined β -bulges²⁰. Some of the more extended interstrand connections are not well ordered, but this occurs at only one of the open ends (the bottom) of the barrel. Electron density for the connection between strands 1 and 4 extends to residue 45 only in subunit B, but the remaining connection to residue 55 is missing in both subunits, and electron density for the connection between strands 5 and 6 (residues 135–155) cannot be interpreted with confidence. The connection between strands 5 and 7 (163–167) is in well-defined electron density, and is the only connection to extend from one open end of the barrel to the other. The interior of the barrel is a hydrophobic core. The active site Ser 132 is located on strand 6 of the β -barrel, and His 63 is located in the connection between strands 3 and 4. *b*, Stereoview of the α -carbon trace of the hCMV protease monomer, subunit B. Residues of the β -barrel, including some of the shorter connections (residues 62–66 and 163–167) are shown in blue, more extended connections (residues 24–45 and 89–127) in red, residues beyond position 174 are green; Ser 132 and His 63 and His 157 are magenta.

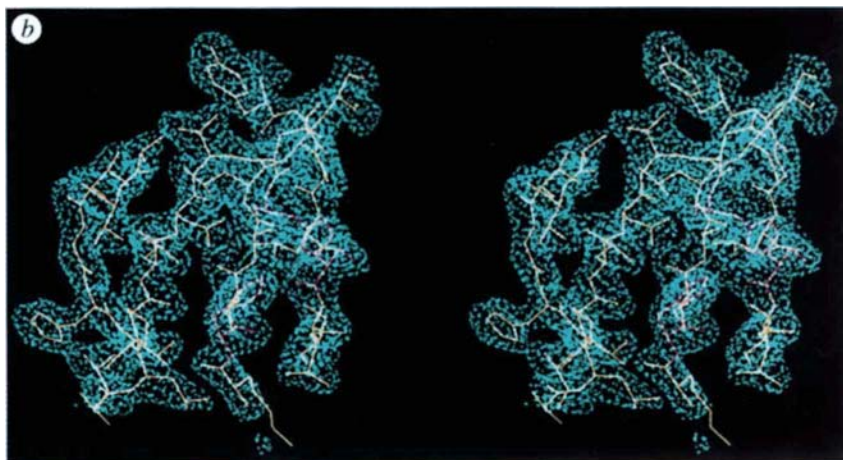
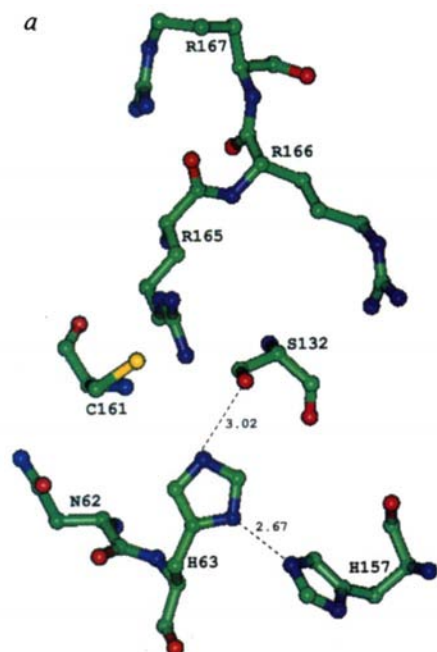
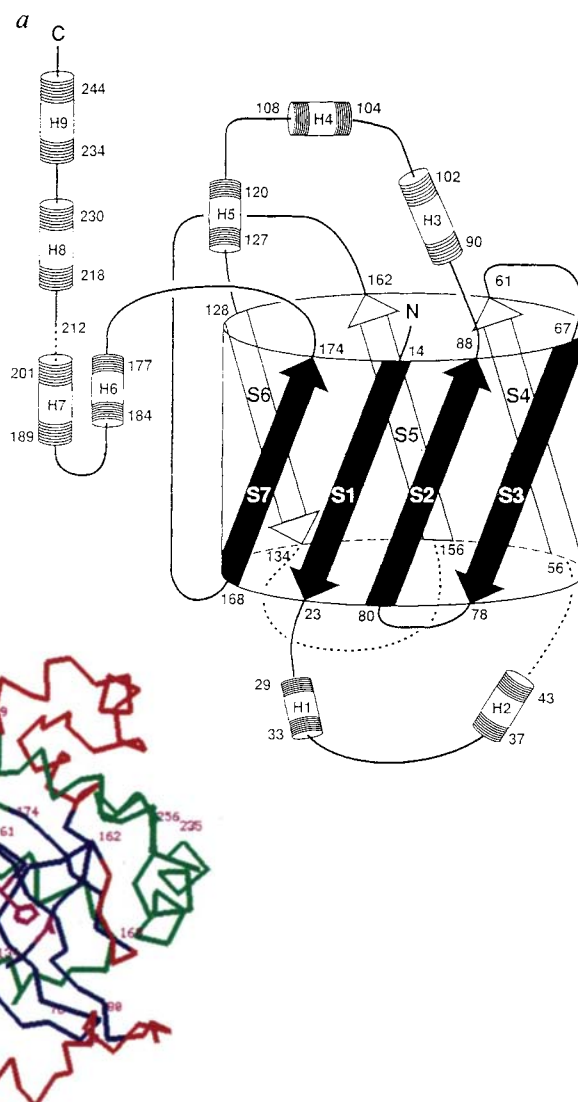


FIG. 2 *a*, The catalytic triad and its surroundings. Hydrogen bonds between the imidazole ring of His 63 and the side chains of Ser 132 and His 157 are shown. The proposed oxyanion hole formed by conserved arginine residues at position 165 and 166 is adjacent to the active-site serine hydroxyl nucleophile. Cys 161 is conserved among viral proteases, but the C161A variant has been shown to be active²⁹. *b*, Stereoview of the electron-density map at the active site. Fourier coefficients for the electron density map are $(2F_o - F_c)$ with phases calculated from the atomic model. The map is contoured at 1.4 σ . Strands 3–6 of the β -barrel are shown. Ser 132, His 63 and His 157 are red.

N-H of trypsin residue 193 with backbone N-H of Arg 165 in hCMV protease, and also indicates that a change in side-chain conformation could align the guanidinium group of Arg 166 with the carboxamide moiety of the Asn 155 side chain of subtilisin in the formation of a herpesvirus protease oxyanion hole. The comparisons indicate that the active site is correctly positioned for attack on the *re*-face of a substrate peptide bond. These correlations with trypsin and subtilisin suggest that, despite the new catalytic triad, the overall catalytic mechanism of herpesvirus proteases is similar to those of previously characterized serine proteases. Unlike other proteases, the binding of anionic intermediates may involve formation of an ion pair with a side-chain guanidinium group. The enhanced strength of this interaction could also contribute to the relatively slow turnover numbers usually reported for the herpesvirus enzymes^{15,16}.

Herpesvirus proteases require high glycerol concentrations for optimal activity¹⁶. It was recently reported that glycerol enhances activity of the protease by promoting dimerization⁶, and the hCMV protease structure is a dimer with local two-fold symmetry. The primary feature of the dimer interface is an α -helix (helix 8, residues 218–230) that runs parallel to the dimer dyad and is recognized by a groove formed by three helices from the other subunit. The groove comprises the equivalent helix 8, as well as helices 3 and 5, which span residues 90–102 and 120–127, respectively. The symmetry positions the two catalytic sites on opposite faces of the dimer (Fig. 3a), and the intersubunit separation between the two sites, that is, the distance between the two Ser 132 residues of the dimer, is 36.5 Å; at this distance, the sites probably function independently.

The structure suggests that dimerization is necessary for activity and that substrate binds to both subunits. Each active site is positioned on a concave surface near the monomer-monomer interface (Fig. 3a). We assume that this surface, with structural components from both subunits, is the binding site for the substrate, assembly protein. This has been investigated using a Connolly surface¹⁷ calculated over the dimeric unit, and the result is shown in

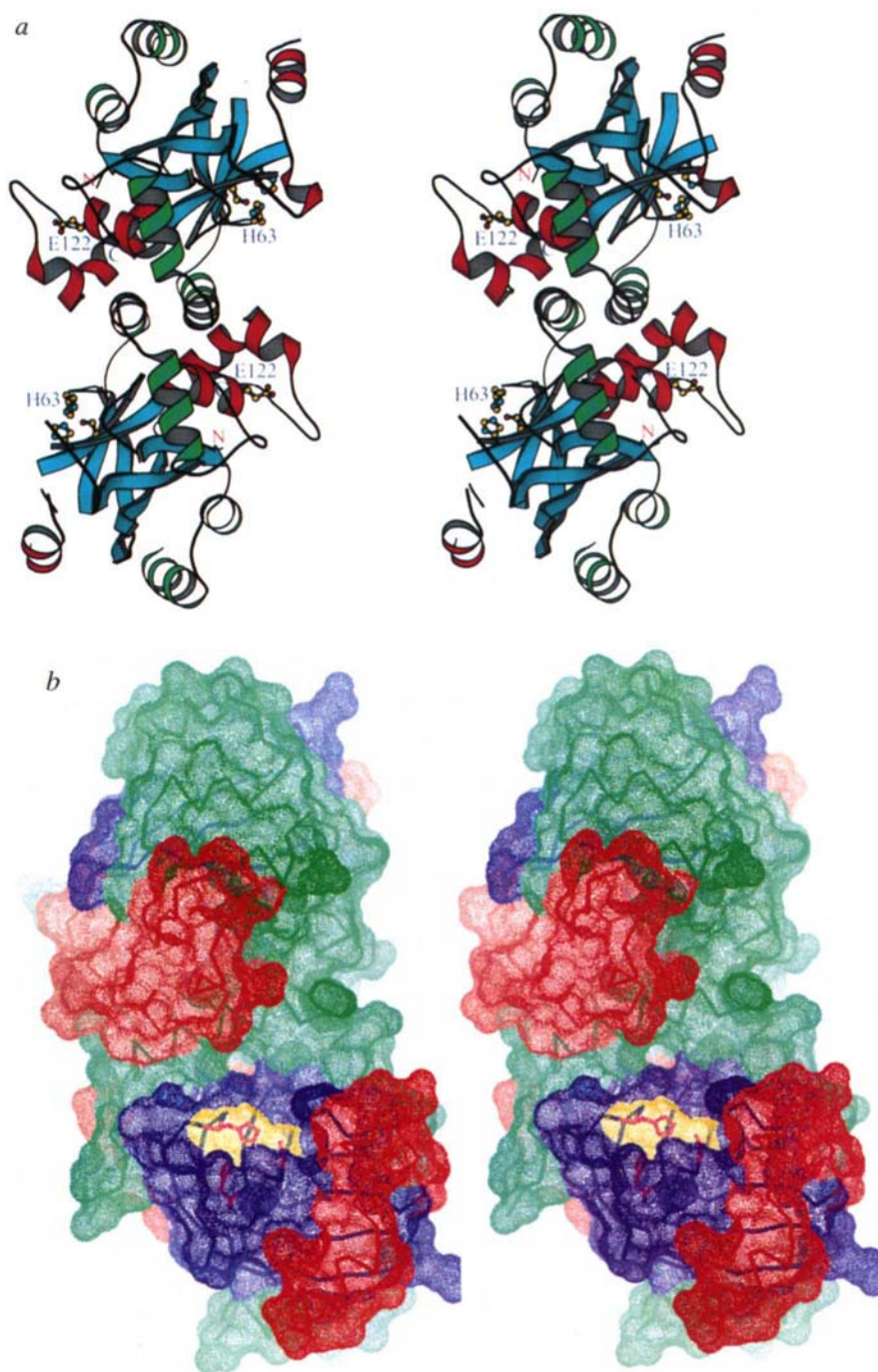


FIG. 3 a, View of the hCMV protease dimer along the local two-fold axis. The active sites are on opposite sides of the dimer. Helices 8, parallel to the dimer symmetry axis, are bound in grooves formed by elements of secondary structure from the opposing subunit. Glu 122 was proposed to be the third member of a putative catalytic triad²⁹. This is inconsistent with the structure, as Glu 122 from neither subunit is located near the active sites. The Glu 122 carboxylate group has an essential structural role in maintaining the protease fold by forming a salt bridge with the Lys 255 side chain near the C terminus; this interaction completes the encirclement of the C-terminal residues around the N-terminal β -barrel domain. Most of the interactions that stabilize dimer formation are hydrophobic; interactions between polar side chains also contribute, but none of the interactions is ionic. Ten potential hydrogen bonds that stabilize dimer formation involve helix 8 and exploit the local dyad symmetry. These are: (from subunit A to subunit B) S 218 OG...S 218 OG, S 225 OG...S 225 OG, D 227 OD1...E 105 N, D 227 OD2...E 106 N, E 105 N...D 227 OD1, E 106 N...D 227 OD1, L 229 O...R 234 NE, R 234 NE...L 229 O, Y 230 OH...G 126 O and G 126 O...Y 230 OH. The surface area of the dimer interface is $\sim 1,405 \text{ \AA}^2$ per subunit. This figure was created with MOLSCRIPT³⁰. b, Stereoview of the Connolly surface of the hCMV protease dimer. The surface was calculated with a probe radius of 1.4 Å over the protease dimer. Colouring is as Fig. 1b except that the surface of the active-site residues of subunit B is yellow. Residues covered by red stippling in the lower right are 24–45 from subunit B. Residues above the active site covered by red stippling are 103–120 from the opposing subunit of the dimer, subunit A. Together, these residues form a groove that partly folds over the active site and is possibly involved in substrate binding.

Fig. 3b, which focuses on one active site. Residues 24–45 from the subunit containing the features site form part of a prominent groove that is completed by residues 102–112 from the opposing subunit. The groove runs by the active site where it appears to fork, indicating that, in addition to residues near the scissile bond, the protease recognizes other structural elements of its substrate.

Alignment of 13 sequences of herpesvirus proteases led to 5 conserved domains (CD1–CD5) being recognized³. Aligned sequences also led to the identification of two additional conserved regions (CD6–CD7)⁸. Comparison of the structure of hCMV protease with these sequences establishes that most of the regions of well-defined secondary structure, including all seven strands of the β -barrel and most of the connections between the β -barrel strands, are in conserved domains. Exceptions include stretches containing residues from the more extended connections, 26–55, 91–127 and 138–154. Helices from the C-terminal domain also overlap with conserved domains: CD6, residues 181–195, overlaps with the two antiparallel helices, H6 and H7; and CD7, residues 221–246, correspond to the remaining two helices, H8 and H9.

The correspondence of the secondary structure with regions of conserved sequences establishes that the folding topology of the hCMV protease monomer is conserved in the proteases from other members of the herpesvirus family. Similar dimerization is also anticipated, as helices 5 and 8 are in conserved domains of the sequence. Residues implicated in substrate binding, including residues 24–43 and 102–112, which form extended strand connections from both open ends of the β -barrel, do not overlap with any of the conserved domains. These residues may therefore confer specificity to herpesvirus proteases in the recognition of their particular substrates.

The targeted design of new antiviral drugs may require the identification of the structural determinants that stabilize the quaternary structure of herpesvirus proteases. The structure identifies a groove on each subunit that binds helix 8 from the opposing monomer. Because we have concluded that dimerization is required for substrate binding and recognition, a molecule that mimics helix 8 would be expected to inhibit protease activity by interfering with the dimerization of the monomers. Similar strategies for rational drug design are being explored in approaches that block the dimerization of retroviral enzymes^{18,19}.

Herpesvirus proteases share no sequence homology with other serine proteases, and so they may form a new subclass¹¹. Three structural classes of serine protease are already known: the chymotrypsin family has a two-domain tertiary structure with each domain containing a six-strand, Greek key, antiparallel β -barrel with very little α -helix; subtilisin has just a single domain with an α/β fold containing a double wound parallel β -sheet²⁰; and wheat serine carboxypeptidase II folds into a single domain with an α/β fold, but contains an extended mixed β -sheet²¹. The hCMV protease structure is rather different, being a dimer of identical subunits that form seven-stranded β -barrel encircled by α -helices. The herpesviruses proteases have a serine protease fold unrelated to those described previously, and provide the fourth example of convergent evolution in the serine protease family of enzymes. □

Methods

Recombinant hCMV protease was expressed and purified¹⁰. Two amino-acid substitutions were introduced at the internal processing site (between residues A143 and A144) using site-directed mutagenesis to prevent autocatalytic formation of two-chain enzyme. This construction, pMON9059, contains V141A and A144P substitutions, and yields greater than 99% one-chain enzyme when expressed and purified from *Escherichia coli*. Crystals of hCMV protease were grown by vapour diffusion from a solution containing 28–33% monomethyl polyethylene glycol 550, 0.05–0.2 M NaCl and 10% glycerol buffered at pH 7.5 with 50 mM HEPES. The crystals are square bipyramids, tetragonal, space group $P4_22_12$ with $a = 76.8$ Å, $c = 172.6$ Å and two protease chains in the asymmetric unit ($V_m = 2.2$ Å³/D). Native data were collected using a MAR image plate with a Rigaku rotating anode X-ray source. A total of 297,289 independent observations of 23,706 unique reflections to a resolution of 2.27 Å

were integrated and scaled using DENZO²² to yield a data set 94.5% complete with an R_{symm} of 0.080. The structure was solved by methods of multiple isomorphous replacement (MIR) with solvent levelling and local averaging. Although six heavy-atom derivatives were used, namely p-hydroxymecuribenzoate (pHMB), ethylmercurithiosalicylate (EMTS), platinum potassium tetrachloride (K_2PtCl_4), trimethyl lead acetate ($(CH_3)_3PbAc$), methylmercurichloride (MeHgCl) and mercuric acetate ($HgAc_2$); there were only three sets of heavy-atom positions. Both K_2PtCl_4 and $(CH_3)_3PbAc$ were bound at the same two sites (set I), pHMB and EMTS were bound to the same four sites (set II), and CH_3HgCl and $HgAc_2$ were bound to the same six sites (set III). The sites in different sets overlap: set II derivatives have the same two heavy-atom sites as the set I plus two extra sites, and set III has the four sites of set II with two additional minor sites. These relationships were used early in the structure determination to establish the local symmetry operator, a dyad axis, relating the two unique protease subunits. All phase calculations were performed using PHASES²³. The initial map, calculated at 2.6 Å resolution with a figure of merit of 0.653, was improved significantly by solvent flattening and local averaging. The final MIR map, after application of solvent levelling and local averaging, had a figure of merit of 0.809. Iterative rounds of model building and refinement used O²⁴ and XPLOR²⁵. Refinement with non-crystallographic restraints was performed using the data with $I > 2\sigma_I$ in the range of 8–2.27 Å resolution. The present model includes 3,155 protein atoms and 18 solvent molecules with r.m.s. deviations from ideality of 0.009 Å in bond lengths, 1.7° in bond angles, and 1.3° in improper angles. The final refinement values are: R_{work} , 0.241; R_{free} , 0.311. The model has many break points. The missing regions are: more than 10 N-terminal residues on both subunits (labelled A and B), and residues 26–55, 135–153 and 200–210 on subunit A, and 42–55, 135–155 and 202–210 on subunit B. Electron density in regions where these residues would be expected is either poorly defined or absent. Attempts to fill the gaps by modelling extra residues increased R_{free} after refinement. These regions of the molecule may be inherently flexible, as they include internal autocleavage sites at residues 143 and 209 (refs 26, 27). During refinement, non-crystallographic symmetry restraints applied to residues 13–25, 56–108, 118–132, 157–191 and 211–247. After refinement, the r.m.s. deviations between equivalent C α atoms in the two subunits are 0.13 Å for the restrained regions, and 0.50 Å for the unrestrained regions. Subunit B, with more defined structure and a lower average temperature factor (23 versus 29 Å²), seems to be better ordered than subunit A. Electron density at the C terminus is well ordered in both subunits; this is probably related to the formation of several salt bridges, such as those between Lys 255 to Glu 122, and Ala 256 to Lys 242 and Glu 249. In both subunits, 88% of the residues in the present structure are in most favoured ϕ – ψ regions, 11% are in additional allowed regions, and 1% are in generously allowed regions as determined by analysis with PROCHECK²⁸. No residues are in disallowed regions. □

Received 20 May; accepted 29 July 1996.

- Welch, A. R., Woods, A. S., McNally, L. M., Cotter, R. J. & Gibson, W. *Proc. Natl Acad. Sci. USA* **88**, 10792–10796 (1991).
- Liu, F. & Roizman, B. *J. Virol.* **65**, 5149–5156 (1991).
- Gibson, W. & Roizman, B. *J. Virol.* **10**, 1044–1052 (1972).
- Preston, V. G., Coates, J. A. V. & Rixon, F. J. *J. Virol.* **45**, 1056–1064 (1983).
- Gibson, W., Welch, A. R. & Hall, M. R. T. *Perspect. Drug Discov. Design* **2**, 413–426 (1994).
- Darke, P. L. et al. *J. Biol. Chem.* **271**, 7445–7449 (1996).
- Margosiak, S. A., Vanderpool, D. L., Sisson, W., Pinko, C. & Kan, C. C. *Biochemistry* **35**, 5300–5307 (1996).
- Loutsch, J. M., Galvin, N. J., Bryant, M. L. & Holwerda, B. C. *Biochem. Biophys. Res. Commun.* **203**, 472–478 (1994).
- Murzin, A. G., Lesk, A. M. & Chothia, C. *J. Mol. Biol.* **236**, 1382–1400 (1994).
- Holwerda, B. C. et al. *J. Biol. Chem.* **269**, 25911–25915 (1994).
- Welch, A. R., McNally, L. M., Hall, M. R. & Gibson, W. *J. Virol.* **67**, 7360–7372 (1993).
- Dilanni, C. L. et al. *J. Biol. Chem.* **269**, 12672–12676 (1994).
- Huber, R., Kukla, D., Steigemann, W., Deisenhofer, J. & Jones, A. J. *Mol. Biol.* **89**, 73–101 (1974).
- Bode, W., Papamokos, E. & Musil, D. *Eur. J. Biochem.* **166**, 673–692 (1987).
- Dilanni, C. L. et al. *J. Biol. Chem.* **268**, 25449–25454 (1993).
- Burck, P. J. et al. *J. Virol.* **68**, 2937–2946 (1994).
- Connolly, M. L. *J. Appl. Crystallogr.* **16**, 548–558 (1983).
- Babe, L. M., Rose, J. & Craik, C. S. *Protein Sci.* **1**, 1244–1253 (1992).
- Huber, R., Baillou, J. G., Rittinger, K., Chermann, J. & Goody, R. S. *J. Biol. Chem.* **270**, 28642–28646 (1995).
- Richardson, J. S. *Adv. Protein Chem.* **34**, 167–339 (1981).
- Liao, D. I. & Remington, S. J. *J. Biol. Chem.* **265**, 6528–6531 (1990).
- Otinowski, Z. in *Proceedings of the CCP4 Study Weekend: Data Collection Processing* (SERC Daresbury Laboratory, Warrington, 1993).
- Furey, W. & Swaminathan, S. *Methods Enzymol.* (in the press).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. *Acta Crystallogr.* **A47**, 110–119 (1991).
- Brünger, A. T., Kuriyan, J. & Karplus, M. *Science* **235**, 458–460 (1987).
- Baum, E. Z. et al. *J. Virol.* **67**, 497–506 (1993).
- Jones, T. R. et al. *J. Virol.* **68**, 3742–3752 (1994).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Cox, G. A., Wakulchik, M., Sassmannshausen, M., Gibson, W. & Villarreal, E. C. *J. Virol.* **69**, 4524–4528 (1995).
- Kraulis, P. J. *J. Appl. Crystallogr.* **24**, 946–950 (1991).

ACKNOWLEDGEMENTS. We thank R. Fletterick for critically reading this manuscript.

CORRESPONDENCE and requests for materials should be addressed to H.-S. Shieh (e-mail: hsshie@ccmail.monsanto.com). Coordinates of the hCMV protease have been deposited in the Protein Data Bank, accession no. ICMV.